THREE ESSAYS ON THE ECONOMICS OF EDUCATION

A Dissertation

by

JAEGEUM LIM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jonathan Meer
Committee Members,	Kalena E. Cortes
	Mark Hoekstra
	Jason M. Lindo
Head of Department,	Timothy J. Gronberg

December 2016

Major Subject: Economics

ABSTRACT

This dissertation investigates, in a causal way, how interactions of student with teachers and peers affect his or her educational outcome.

First, we use random assignment of students to Korean middle school classrooms and show that female students perform substantially better on standardized tests when assigned to female teachers; there is little effect on male students. We find evidence that teacher behavior drives the increase in female student achievement.

Also, we shed light on the importance of teacher student gender matches in closing the gender gap, especially in STEM fields in the long run. We exploit data from middle schools in Seoul, South Korea, where students are randomly assigned to a middle school and where students and teachers are randomly assigned to a physical classroom. Our finding is that female students taught by a female versus a male teacher keep achieving higher scores in standardized tests compared to male students even four years after the exposure to the teacher. We also find that if female students learn math from female teacher in seventh grade, then the likelihood increases that they take higher-level math courses and aspire to a STEM degree in their 11th grade. We show the evidence that the long lasting gender gap effects are driven by student's behavioral change.

Lastly, we examine classroom peer effects on BMI. In response to increasing child obesity, many researchers have studied the sources of obesity, with social scientists focusing on peer effects. However, three well-known challenges make it difficult to find peer effects. We avoid self-selection using random assignment of classroom peers. To address common environmental factors and reflection problem, we instrument for peer BMI with number of peer siblings. We find that if peer BMI increases by one unit, student's own BMI increases by 0.83 units and that the reduced social outdoor activities drive the effect.

DEDICATION

To my family, my teachers, and my friends

ACKNOWLEDGMENTS

I am deeply grateful to Dr. Jonathan Meer. He selected me as his student when I was in dire need of help. I will not forget the moment when he said, “Let’s discuss plans for this year.” To the finishing line to my doctoral degree, he ran together with me. Whenever I met him, he had the best advice, which always led me to break through in the research. I was slow in finding an answer, but he put his trust in me, and waited and waited. He was a guardian in my academic life in the United States, and I really appreciate his care.

I also would like to express appreciation to my committee members. Thanks to Dr. Mark Hoekstra, I chose public economics as my field of interest. He has always encouraged me in researching and showed me how to highlight the good aspects of my research. I benefited greatly from his detailed comments on our papers, which were much better than referee reports. I also would like to thank Dr. Jason Lindo and Dr. Kalena Cortes for their invaluable comments and suggestions. It was very fortunate for me to have advice from professors as well-versed in microeconometrics and education literature as them.

I gratefully acknowledge Dr. Hae-Shin Hwang. His mentorship during the first year of doctoral courses enabled me to endure the hardship and to recover from the failure. I regret that I failed to keep the promise that I made him to graduate within three and a half years, but thanks to the promise, I was able to come back to normal doctoral student’s life whenever I was slack in studying. A special thanks goes out to Dr. Jaeun Shin and Dr. Youngsub Chun. With their help, I could start to study again. Their belief in my potential was one of the sources that I kept motivated.

I would like to thank especially María del Socorro Padilla-Romo, one of the best peers. She was superior to me in every aspect, and helped me much. She was as good as a tutor for me during the first year, and afterwards showed me the exemplary path to grow as a researcher. I also thank to Korean students, senior or junior to me: Jonghoon Kim, Sung-Ick Cho, Sungmin Han, Jinkook Lee, Wongi Kim, Jin Hyung Lee, Douyoung Lee, and Eunseong Ma. Their tips on studying and information on living in the U.S. helped me much whenever I was in trouble. Also, whenever I came up with a research idea, they played the devil's advocate for me.

My gratitude also goes to Jeongkyu Kim for Project Texas he set up for my research to be successful. I also deeply appreciate Yonghoon Choi, Jooah Song, Jeonghee Shim, Younghoon Koh, Seongo Cho, Sungmin Kang, and Jaeyoung Park for helping me to contact education officials or to obtain detailed knowledge and data on Korean education.

Special thanks are reserved for my family. Talking to my parents, Hoesub Lim and Jeongja Cho made me always comfortable and free of any worries. Whenever I was exhausted, surprisingly, my mother knew it, and thanks to her tips to stay healthy I regained strength to study. My father, also a former teacher in secondary school, motivated me to study teacher effects, and his experience enhanced my understanding of Korean education data. My parents in law, Kyungwon Han and Hee Kim also made me feel easy and encouraged my progress. Especially, my mother in law prayed for my success even when she was seriously ill. With continuous support and encouragement from my sister, brother, and brother in law, Youjeong Lim, Jaeyeon Lim, and Kyul Han, I was always reassured. My wife, Song Yee made tasty and nutritious lunch and dinner to bring school every day during my first year in Ph.D., which made me look so healthy that I was considered not studying hard. From time to time she made questions about my research, which were useful in clarifying the reasoning in the research. But, above all, her staying with me has been the greatest help. My son, Donghyun also is one of the greatest contributors to this dissertation. He taught me English that he learned from his day care and kindergarten. Spending time with him was much fun and refreshing most of the time, though not always. Thank you for waiting for a long time.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by an advisory committee consisting of Professors Jonathan Meer, Mark Hoekstra, and Jason M. Lindo of the Department of Economics, and Professor Kalena E. Cortes of the Department of Public Service and Administration.

All work for the dissertation was coauthored by Professor Jonathan Meer and the student. Section 2 is forthcoming in the Journal of Human Resources.

Funding Sources

Graduate study was supported by fellowships from Korean National Assembly and Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE | Page

LIST OF TABLES

1. INTRODUCTION

The dissertation, which comprises three articles, examines the effects of a student's interactions with his or her teachers and peers on educational outcome. The first and second articles investigate the short and longer term impacts of teacher student gender matches on academic achievement, and the third one deals with peer effects on student's body mass index. To find the causal relationship, we use unique feature in secondary education in South Korea: random assignment of students to a *physical* classroom where students stay and each subject teacher rotates to give them a lesson.

1.1 Short-term Impact of Teacher-Student Gender Matches

Over the past 40 years gender gap in academic performance have persisted. In response, many studies have tried to explain the sources of the gender gap and proposed the measures to close it. Among them are single-sex schooling and teacher-student gender matching.[1] We focus on the latter.

One of the biggest empirical challenges that researchers face in finding the causal relationship in this literature is nonrandom sorting of students. For example, if a female teacher is more likely to be assigned to high achieving students, estimating the teacher effect without controlling for student's unobserved characteristics would suffer from selection bias. One standard way to tackle with this problem has been to use student fixed effects. However, even this approach can lead to biased estimates if teacher-student matching is done systematically in a way that affects student's academic achievement. If a principal matches a female teacher with female students of high propensity to achieve, the same

[1]For single-sex schooling, see Jackson (2012); Park, Behrman and Choi (2013); and Lee et al. (2014), and for teacher-student gender matching, see Dee (2007); Holmlund and Sund (2008); Hoffmann and Oreopoulos (2009); Carrell, Page and West (2010); Cho (2012); Fairlie, Hoffmann and Oreopoulos (2014); Paredes (2014); Antecol, Eren and Ozbeklik (2015); and Muralidharan and Sheth (2016).

gender teacher effect estimates can be biased even including both student fixed effects and teacher fixed effects.

We circumvent this problem by exploiting the fact that middle school students are randomly assigned to a physical classroom in which they stay throughout a school day and which each subject teacher visits for a lesson. Also, we use the fact that teacher assignment to the classrooms is done irrespective of student's and teacher's characteristics. Combining the two facts on the classroom assignment in Korean middle school, we obtain exogenous variation in teacher-student gender matching, which enables us to find a causal link between the gender matching and student's academic performance.

Using data from Korean Educational Development Institute in 2004 (KEDI 2004), we show in various ways the evidence that random assignment is in place: first, we survey 198 schools in our sample on the student and teacher assignment. Then, we find evidence that the teacher-student matches are independent from student's and teacher's observable characteristics using resampling technique which Carrell and West (2010) use to test if the algorithm that students are placed into course sections at the United States Air Force Academy (USAFA) looks to be random. Similarly, a series of Pearson's χ^2 tests are done to check for the independence of students' characteristics and their assigned classroom. Also, we show teacher characteristics look similar between female and male students. Finally, we show the stability of coefficients across the specifications, even including both student and teacher fixed effects.

The main result we find is that if a male teacher is switched to a female teacher, the gender gap in academic achievement between female and male students increase by 0.1 standard deviations. Recalling Carlsson et al. (2015)'s finding that 0.01 standard deviations of achievement amounts to 10 days of schooling, the effect is sizable. We also find the gender gap effect is comprised of 8% of a standard deviation increase in female's performance and insignificant decrease in male's achievement, suggesting that gender matching policy would benefit female students without hurting male students' performance.

Finally, we find the mechanism behind the positive gender gap effects. While Dee (2007) finds a student centered mechanism where a student is more engaged in study when taught by the same gender teacher, we find the evidence on teacher centered mechanism; we find that female students are more likely to report that female teachers encourage them to express themselves and give them equal opportunity to participate.

1.2 Persistent Effects of Teacher-Student Gender Matches

Female representation in science, technology, engineering, and math (STEM) occupation has been low (Corbett and Hill, 2015). It is important to understand the reason because large portion of gender pay gap is explained by gender gap in STEM careers (Brown and Corcoran, 1997). We shed light on this issue by focusing on the lasting effects of teacher-student gender interactions in seventh grade.

To date, only a few studies (e.g. Carrell, Page and West, 2010 and Lavy and Sand, 2015) have examined the longer term effects of the role of teacher or instructor gender, mainly because of the data availability; student's or parent's influence over teacher assignment hinders researchers from identifying the pure teacher/instructor gender effects. In contrast, with the help of random assignment of students into a classroom at middle schools in South Korea, together with our longitudinal data set that tracks seventh grade students until their 11th grade, we can distinguish the longer term effects of teacher-student gender matches with confounding factors. Admittedly, ability group practice in some schools might bias our estimates. We address this issue by comparing students within a school by subject by ability group level cell.

To check whether we have random variation in our data, Seoul Education Longitudinal Survey of 2010 (SELS2010), we conduct various tests. We begin by using resampling technique as in the first paper to show the evidence that students are randomly placed to a classroom, and the subject teachers' characteristics are independent from the student's past test scores in the classroom. Next, we investigate whether the characteristics of students who are taught by a female versus a male teacher look different. We also compare teacher characteristics between female and male students. Importantly, we show female teacher gender in seventh grade is not correlated with student's characteristics. The stability of the contemporaneous teacher-student gender interaction effects across the various specifications reconfirms that we have exogenous variation in teacher-student gender matching. We also find that the magnitude of gender gap effect in seventh grade (0.14 standard deviations) is similar to that in ninth grade (0.10), which we find in the first paper. This implies that the random assignment of students to a classroom in South Korea is in place.

Our main finding on longer term effects is that gender gap effect in seventh grade does not fade even four years after the exposure to the teacher; female students would perform better in eighth through 11th grades relative to their male counterparts if a male teacher in seventh grade is switched to a female teacher. The magnitudes of the gender gap effects range from 0.10 to 0.16 standard deviations by the grade. We show that the estimates do not suffer from attrition bias by regressing the likelihood of attrition in each grade on our variables of interest (i.e. student gender, teacher gender, and their interaction). We also find that the gender gap effects are significant and positive on student's advanced course-taking and plan to seek STEM degree.

To find the possible mechanisms behind the persistent effects, we first test whether having a female teacher in a grade is correlated with our variables of interest. We also show that seventh grade teacher-student gender matching does not influence the probability of being in high ability group in later years. However, we find the significant positive gender gap effects on student's going to high school of higher quality in terms of teacher value added, classmates' quality, and the quality of the former students in the high school. This result, combined with the results on high school choice, weighs in favor of the student centered mechanism, whereas we find the teacher centered mechanism for contemporaneous effects in the first article.

1.3 Influence of Classroom Peers on BMI

The ratio of 15-year-old children in OECD countries reporting to be overweight has steadily risen since 2000 (OECD, 2015*b*). The fact that obesity has increased at all income levels in the United States (Chang and Lauderdale, 2005) highlights the importance of social factors rather than individual characteristics for the explanation. Because adolescents are heavily influenced by their peers, peer effects may play a role.

However, empirical research on peer effects has been difficult because of the well-known issues such as self-selection, common environmental factors, and reflection problems (Manski, 1993; Epple and Romano, 2011). We overcome self-selection problem using random assignment of students into a classroom at a middle school in South Korea. Because Korean middle school students take courses in the same classroom for a day throughout a school year, classroom peers are appropriate social network to be examined. To address reflection problem, we use peers' mean number of siblings as an instrumental variable for peers' average BMI; a number of studies (e.g. Hesketh et al., 2007; Chen and Escarce, 2010; Haugaard et al., 2013; and de Oliveira Meller et al., 2015) find number of siblings is highly correlated with child's BMI and likelihood of being obese, as is shown in our data. Also, arguably peers' average number of siblings cannot directly affect student's own health condition; we show that peers' mean number of siblings is not correlated with student's characteristics.

We use seventh grade data of Gyeonggi Education Panel Study of 2012 (GEPS2012), which surveyed seventh graders in 63 middle schools in Gyeonggi province that surrounds Seoul, South Korea. First, we conduct a series of Pearson's χ^2 tests for independence of various students' characteristics (e.g. student's gender, number of siblings, father's and mother's education, parents' marital status, as well as whether parents own their own home) and their assigned classroom. 34 of 1,111 available p-values (3.1%) reject the null hypothesis of independence at 5 percent significance level, implying the random classroom assignment. Also, we show that student and teacher characteristics are not correlated with peers' BMI of their seventh grade.

Naïve OLS result shows that when peers' average BMI increases by one unit, student's BMI decreases by 0.74 units. The negative coefficient reflects that students' BMIs are balanced across classrooms. IV estimate indicates a one unit increase in peers' mean BMI would increase student's BMI by 0.83 units. Also, the effect of seventh grade peers' mean BMI on student's BMI in eighth grade is still significant and positive (0.56). The contemporaneous and longer term effects of seventh grade peers' mean BMI are stable regardless of including student's, peers', and teacher's characteristics.

2. THE IMPACT OF TEACHER-STUDENT GENDER MATCHES RANDOM ASSIGNMENT EVIDENCE FROM SOUTH KOREA*

2.1 Introduction

Gender gaps in academic performance, with girls generally outperforming boys in language arts and boys generally outperforming girls in math, have persisted despite decades of effort to close them (OECD, 2015*a*). Understanding the causes of these gaps is crucial, especially at younger ages, as they may lead to gender differences in later course-taking, occupational choices, and labor market outcomes (Lavy and Sand, 2015).

One possible source of gender-based disparities is whether a student and a teacher share the same gender. These gender interactions may affect academic performance through changes in the behavior of both parties, that is, through student- or teacher-centered mechanisms. Role-model effects, an example of the former, predict that students will be more engaged in study when they are taught by the same-gender teacher (Dee, 2007). As an example of the latter, a teachers might assign less difficult homework questions to girls if he or she believes that girls are less capable in math than boys (Jones and Dindia, 2004).

The primary threat to identifying the causal effect of teacher-student gender matches is the nonrandom sorting of students that typifies classroom assignment in most contexts. For instance, students with a lower propensity to achieve academically may be more likely to be assigned to a teacher of a particular gender. Beginning with Dee (2007), the standard approach in this literature has been to use student interactions with multiple teachers across different subjects. By using estimates including student fixed effects, unobserved student characteristics that are correlated with student quality and teacher gender will not

*Part of this section is reproduced courtesy of the University of Wisconsin Press from Lim, Jaegeum and Jonathan Meer. "The Impact of Student-Teacher Gender Matches: Random Assignment Evidence from South Korea." *Journal of Human Resources* 52.4 (2017). © 2017 by the Board of Regents of the University of Wisconsin System.

bias estimation. Dee uses the fact that the National Education Longitudinal Survey of 1988 surveys two teachers for every student to estimate within-student teacher-gender effects; he finds evidence of substantial positive impacts on academic achievement of being assigned to a teacher of the same gender. Moreover, Dee uses subjective evaluations of both teacher and student perceptions to show that students are less likely to be seen as disruptive when evaluated by a teacher of the same gender, and more likely to report interest in that academic subject. Using a different approach, Muralidharan and Sheth (2016) exploit panel data from India – in particular, schools with only one classroom per grade, in which there can be no sorting of students. They find that female primary school students perform significantly better with female teachers, with no impact of teacher gender on male students.

On the other hand, several studies find no effect of teacher gender. Holmlund and Sund (2008) use Swedish secondary-school panel data and identify the impact of same-gender teachers using teacher turnover. Once they control for subject-specific gender effects, they find no impact of gender matching on student performance. Cho (2012) uses math and science test score data from 15 OECD countries and shows that there is no significant effect of teacher-student gender matching in eight of these countries, including the United States. Most recently, Valentina Paredes (2014) examines role model and teacher bias effects with data from Chile, finding small but statistically significant gender-matching effects for girls and no effects for boys, as well as suggestive evidence that role model effects drive the result.

However, this within-student estimation approach – even when including *teacher* fixed effects – is insufficient if students and teachers are systematically matched on characteristics correlated with gender. For instance, suppose female students who would benefit relatively more from having a female teacher are more likely to be assigned to female teachers who, themselves, are better role models for female students. In this case, a positive student-teacher gender interaction effect reflects sorting. As Dee notes, “the internal validity of such within-student comparisons could still be compromised by the nonrandom sorting by students with subject-specific propensities for achievement and by unobserved teacher and classroom traits correlated with gender.” He finds some such evidence in the NELS:88 with a number of indirect tests, particularly in the assignment of female math teachers. Other studies lacking random assignment must indirectly show that the identification strategy holds; for instance, Valentina Paredes (2014) uses previous-year’s test scores to control for achievement propensity.

For identification, we exploit a unique feature of secondary education in South Korea: the random assignment of students into a classroom, where students remain throughout the school day. We provide evidence for our identifying assumption in a number of ways: first, as-good-as-random assignment of students to classrooms is a strict policy in South Korea. We confirm that schools follow this policy by surveying a large number of them on the topic. We also show that assignment to classrooms within a school is uncorrelated with observable characteristics. Furthermore, students who are assigned to same- and opposite-gender teachers look similar in their observable characteristics. Finally, our results do not differ when additional controls, student fixed effects, or teacher fixed effects are included, as one would expect if assignment is truly random.

Our reliance on random assignment obviates potential sorting issues that have been a major concern in previous work. In this way, our approach is most similar to two previous papers. Antecol, Eren and Ozbeklik (2015) exploit the random assignment of students in an experiment testing the efficacy of Teach for America, a program that trains and places high-achieving new teachers at disadvantaged schools. They find that female elementary school students with female teachers perform *worse* than those with male teachers. However, this negative effect disappears for female teachers with stronger math backgrounds. At the higher education level, Carrell, Page and West (2010) use random assignment of cadets at the United States Air Force Academy to compulsory math and science courses and show that female professors significantly reduce the gender gap in performance for female students.[2]

Our study is also unique in that we provide more recent evidence from an age group similar to that studied in Dee (2007) and, importantly, our empirical setting is a culture with somewhat different gender norms than many of those previously studied. South Korea is ranked 39^{th} of 57 countries in its residents' attitudes towards gender equality, much lower than the countries studied in the analyses above: 29^{th} for Chile, 18^{th} for the United States, and 2^{nd} for Sweden (Brandt, 2011).[3]

Results show that female students' performance is positively influenced by having a female teacher, but that there is little same-gender teacher effect for males. The overall increase in the female-male performance gap of about a tenth of a standard deviation is comparable in size to those found by Dee (2007) and Carrell, Page and West (2010). Unlike our findings, though, both of those papers find that the impact is divided about evenly between reduced performance by males and increased performance by females. Our effect is similar in magnitude to an increase of one standard deviation in teacher quality (Chetty, Friedman and Rockoff, 2014).

The impacts are primarily concentrated in mathematics and English language scores, as compared to Korean language scores. We also provide some suggestive evidence that teacher-centered mechanisms are behind these impacts, with female students reporting that their female teachers are more likely to encourage them and to give them an equal opportunity to express themselves.

[2] Other evidence on gender-matching effects on student grades, course-taking, and persistence in colleges is mixed; see, for example, Canes and Rosen (1995), Bettinger and Long (2005), and Hoffmann and Oreopoulos (2009).

[3] Our work is also related to the literature on the impact of single-sex schools. Park, Behrman and Choi (2013) find significant positive impacts of single-sex schooling using random assignment in South Korea, while Jackson (2012) exploits the nature of rules-based school assignment in Trinidad and Tobago and finds little effect for most students.

2.2 Data

We use cross-sectional data collected by the Korean Educational Development Institute (KEDI) in July 2004, at the end of the first semester of middle school in South Korea. The target schools, covering 6.8 percent of the relevant population in South Korea in 2004, were selected by proportionate stratified random sampling. Our initial sample consists of 197 schools, 777 Korean, English, and mathematics teachers linked to surveyed classrooms, 14,372 students, and 11,944 parents. Thirty-five of the schools had all-female students and 35 were all-male; 84 classrooms are single-sex within 127 coed schools.[4] Restricted-use data provided by KEDI allows us to link students to classrooms.

In addition to an extensive set of questions, students' responses were linked to their scores on the Student Achievement Test administered by the Seoul Metropolitan Office of Education (SMOE). Students in the sample were tested at the beginning of the second semester of ninth grade in three courses: Korean language, English language, and mathematics; 12,363 students' test results were collected.[5]

The teacher questionnaire includes information on teachers' classroom assignments, which we use to link students with their subject teachers. Beginning with 37,034 student-subject combinations with test score information, we first drop 6,033 observations without classroom or teacher information. Of these, 42 observations from 14 students have missing classroom information and 5,991 observations from 224 classrooms do not have teacher information due to nonresponse by teachers, reducing the number of teachers in the sample to 777 and the number of students to 12,305.[6] For our primary sample, we also drop 6,442

[4]As discussed below, excluding single-sex schools or single-sex classrooms does not change our results.

[5]This exam is administered to 9^{th} graders in Seoul every September; these students would have taken the test regardless. Students living outside of Seoul but in the KEDI sample took the same exam on the same day.

[6]A concern is that teacher non-response could somehow be correlated with their impact on students of different genders. While we cannot completely exclude this possibility, students dropped from the sample due to teacher non-response have similar test scores ($p = 0.77$) as those remaining in the sample. There were also no statistically significant differences in the other student characteristics we examined.

observations for students with multiple subject teachers, for which we could not make a student-teacher match representing just one student and one teacher. We also show results including these observations, which are unchanged from those excluding them. This results in 24,489 student-teacher pairings representing 11,659 students and 502 teachers. Among them, 33 percent of observations correspond to a female student with a female teacher, 16 percent are a female student with a male teacher, 32 percent are a male student with a female teacher, and the remaining 19 percent are male students with male teachers.

2.2.1 Student Assignment in South Korea

Elementary school graduates in South Korea are randomly assigned to middle schools within their district.[7] At the beginning of each academic year (March 1^{st}), middle school students in South Korea are assigned a classroom where they remain throughout a school day, and where each subject teacher visits to present a lesson. Be it private or public, schools in South Korea use some form of random assignment to classrooms due to both strong social norms and government policies (Kang, 2007). The most common approach is to order students by their academic performance in the previous year and assign them across classrooms. As an example, the top ranked student would be assigned to the first classroom, the second-ranked student assigned to the second classroom, and so on.[8] To confirm this point, we surveyed local Offices of Education on schools' rules for classroom assignment for the 197 schools in our sample.[9] All but one of the 180 responding schools

[7]Since 1996, students in districts whose superintendents allow it are permitted to list several preferred schools. They are entered into a lottery for each school on their preferred list (Korea Legislation Research Institute, 2011).

[8]Kang (2007) uses this same random assignment feature and a different data set on the performance of Korean students to examine peer effects. As mentioned above, Park, Behrman and Choi (2013) examine the effect of single-sex education on college-going behavior using data from Seoul, in which students are not allowed to list preferred schools. Lee et al. (2014) examine schools in the Seoul metropolitan area to study the effects of single-sex versus co-educational schooling on academic performance.

[9]Note, of course, that schools were responding eleven years after the KEDI survey was conducted. In recent years, the Korean education system has shifted somewhat from its original strictly egalitarian approach, so it seems quite likely that these as-good-as-random practices were in place in 2004. See, for example, Byun and Kim (2010), who discuss increased use of ability tracking in South Korea over the past

with more than one classroom per grade reported that they used this method of classroom assignment, with the sole exception being a school that used alphabetical order of names to assign students.

2.2.2 Teacher Assignment in South Korea

Even with random assignment of students to classroom units, the internal validity of our approach is threatened if teachers are systematically assigned to those classrooms in a way that is related to their gender. For example, female teachers might be assigned to classrooms that, by chance, have students with less-involved parents. There are no written government guidelines on teacher assignment; we interviewed a number of current teachers and principals to gain insight into the process. First, homeroom teachers are assigned, either by lottery or a committee, to a particular classroom. These teachers, who teach a subject themselves, are responsible for discipline, taking attendance at the start of the day, and overseeing study halls before and after school. Subject teachers' classroom assignments are generally determined in an ad hoc way that is unrelated to student or teacher characteristics. For example, one subject teacher may take odd-numbered classrooms while the other takes even-numbered ones. We surveyed the schools in our sample on these policies as well, with 141 of 153 responding schools reporting that they assign subject teachers without considering student or teacher characteristics. The remaining 12 schools reported considering teachers' characteristics, such as experience, in making the assignment; our results are unchanged by excluding these schools, and we once again note that we conducted our survey eleven years after our data were collected. In Subsection 2.3.1, we further examine whether random assignment holds in our data based on students' and teachers' observable characteristics.

---

decade.

2.3 Methodology

2.3.1 Tests of Random Assignment

While the institutional setting we study is clear that students are randomized across classrooms without respect to teacher gender, we also provide empirical evidence to support our identification strategy. We begin by following Carrell and West (2010), Lehmann and Romano (2005), and Good (2006) in using resampling techniques to test the randomness of teacher and student matching in terms of student's observable characteristics. First, for each classroom within a school, we randomly draw 10,000 synthetic classrooms of the same size from the sample of all students in the school, without replacement. We do so for each of the three subjects – Korean, English, and mathematics – and for each of six variables (indicator variables for student being male, parents being married, father with BA degree or higher, mother with BA degree or higher, having housing ownership, and having Internet access at home). Then, for each subject and characteristic combination, and for each classroom within a school, we calculate the number of students with the characteristic within a classroom.[10] We obtain an empirical p-value, namely, the proportion of the 10,000 resampled classrooms with fewer students with the characteristic (for example, male) within the observed classroom.

Under random assignment, any p-value will be observed with equal probability; we therefore expect the empirical p-values to be uniformly distributed. We test whether the distribution of the empirical p-values for each subject and characteristic combination is uniform using Kolmogorov-Smirnov and χ^2 goodness-of-fit tests. Table 1A presents the results of this exercise, aggregating results over school subject for brevity. Overall, we reject random assignment for 34 out of 1,942 (1.8 percent) school-by-subject-

[10]Carrell and West (2010) use the sums of SAT scores or academic composite to obtain empirical p-values. Similarly, we sum the indicator variables.

Table 2.1: Randomness Check

	(1) Male	(2) Married Parents	(3) Dad BA Degree	(4) Mom BA Degree	(5) Owning Home	(6) Internet Home
A. Test for Student Assignment						
Kolmogorov-Smirnov test (No. failed / total tests)	7/229	18/343	0/343	0/341	9/349	0/336
χ^2 goodness of fit test (No. failed / total tests)	9/229	16/343	9/343	14/341	15/349	8/336
B. Test for Teacher Assignment						
Female Teacher	0.018	−0.032	0.004	0.033	−0.016	−0.010
	(0.013)	(0.018)	(0.019)	(0.023)	(0.016)	(0.015)
Teacher Experience	0.000	−0.001	0.002	0.000	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Graduate Degree	−0.014	0.014	0.029	0.045*	0.028	0.026
	(0.020)	(0.017)	(0.019)	(0.019)	(0.019)	(0.017)
Teachers College Graduate	−0.019	−0.012	0.000	0.008	0.003	−0.013
	(0.020)	(0.014)	(0.014)	(0.014)	(0.016)	(0.015)
Teachers' Union Member	−0.008	0.014	−0.012	−0.000	−0.000	0.018
	(0.026)	(0.018)	(0.023)	(0.025)	(0.017)	(0.019)
N	530	856	848	720	909	671
R^2	0.210	0.184	0.064	0.221	0.053	0.277

Notes: Each column represents a separate regression. The dependent variable is the empirical p-value from the resampling described in the text, for student characteristics, which are male student, having married parents, having father with BA degree or higher, having mother with BA degree or higher, owning their home, and having access to Internet home in Columns 1 through 6, respectively. Independent variables are teacher gender, teacher experience measured in years, dummies for graduate degree, teachers college graduate, and teachers' union member. Each specification controls for subject and school fixed effects. Standard errors in parentheses are clustered at school level. The Kolmogorov-Smirnov and χ^2 goodness of fit test results indicate the number of tests of the uniformity of the distribution of p-values that failed at the 5 percent level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

by-characteristic test statistics[11] at 5 percent level in the Kolmogorov-Smirnov test and 71 of 1,942 (3.7 percent) test statistics using the χ^2 goodness of fit test. Therefore, we do not find evidence of nonrandom assignment of students into classrooms by observable characteristics.

We also check the random assignment of teachers with respect to student's observ-

[11]For some school by subject by characteristic combinations, test statistics cannot be calculated because of missing variables, only a single classroom from the school remaining in the sample, or the school itself being single-sex for the student gender characteristic.

able characteristics. For each characteristic, we regress the empirical p-values on a set of teacher characteristics, controlling for subject and school fixed effects. The results, in Table 1B, show that only one of thirty coefficients is statistically significant at the 5 percent level. Therefore, there is little evidence of nonrandom assignment of students into classroom with respect to student's observable characteristics.

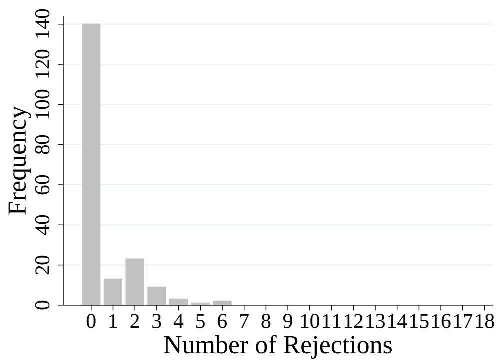
We next turn to testing random assignment with respect to observable characteristics by conducting a series of Pearson's χ^2 tests for independence of a variety of characteristics and the classroom to which they are assigned. Tested characteristics include student's gender, parents' marital status, parents' education, as well as whether parents own their own home and whether student's home has access to the Internet, as proxies for family resources. Parents' education has seven categories and the other variables are indicator variables.

We perform 2,082 Pearson's χ^2 tests across six characteristics and 453 school-subject combinations.[12] We find that 208 (9.99 percent), 115 (5.5 percent), and 38 (1.8 percent) of these p-values are lower than or equal to 10 percent, 5 percent, and 1 percent, respectively. This provides further evidence for the random assignment mechanism described in Subsection 2.2.1.

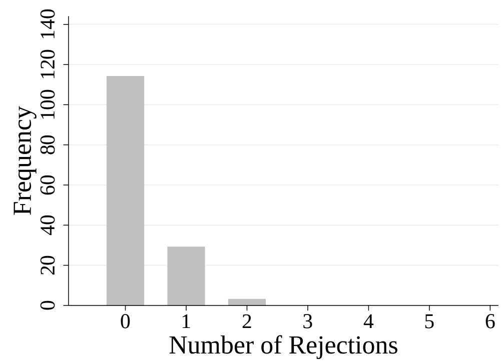
To check whether the rejections are concentrated in particular schools, we examine distributions of the number of rejections by school. Figure 1 shows the distributions for all subjects and each subject. Two schools have a total of six rejections in all subjects combined and one school has five rejections. Only one school has as many as three rejections in one subject, suggesting that no schools that are failing to comply with the random assignment of students to classrooms. Further, omitting the three schools with five or six total rejections from our estimates does not affect the results.

[12]Some combinations cannot be tested due to missing variables, only a single classroom from the school remaining in the sample, or the school itself being single-sex.

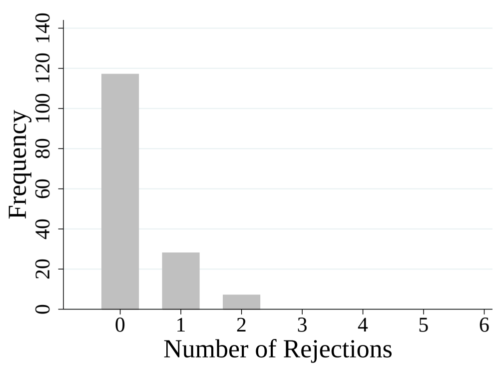
Figure 2.1: Number of Rejections for the Independence by School



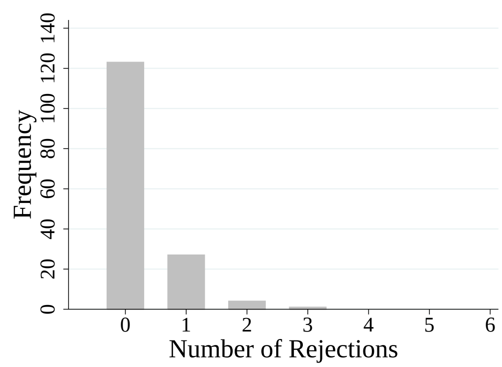
(a) All Subjects



(b) Korean Language



(c) English Language



(d) Mathematics

Another approach is to compare the groups of students taught by same- and opposite-gender teachers. If the students are randomly assigned to the teachers of the same and opposite gender, then the two groups should look similar in terms of observable characteristics.

Table 2.2 presents sample means from our data, with each observation as a student-teacher pair. Recall that the randomization in our sample is *within* schools, though even when looking across schools, the results are fairly well-balanced. In Panel A, the characteristics for female and male students are presented separately by teacher gender, demonstrating that students are not more likely to be assigned to a teacher of the same gender based on observable characteristics. For male students, there is a statistically significant difference for home ownership, but it is economically small. Moreover, since random assignment was done within schools, adjusting for school fixed effects eliminates the significance of this difference. We also show the mean standardized test scores by group as a preview of our results. Female students perform substantially better than male students overall, but particularly when they have female teachers. Meanwhile, male students are not greatly affected by the gender of their teacher. In Panel B, we compare teachers' characteristics when assigned female and male students. As in most schools around the world, female teachers are much more prevalent in our sample, but there are no significant differences in the types of teachers assigned to students of different gender. These results further show that students and teachers are randomly assigned to classrooms irrespective of gender matches.[13]

[13] We also follow Carrell, Page and West (2010) in examining whether the student characteristics in Table 1 predict teacher gender; they are jointly insignificant at $p = 0.31$.

Table 2.2: Comparison of Mean Characteristics

A. Student Characteristics	Female Students				Male Students			
	with FT	with MT	P-value	Observations	with FT	with MT	P-value	Observations
Married Parent	0.903	0.899	0.673	10,326	0.901	0.893	0.315	10,059
	(0.007)	(0.006)			(0.006)	(0.007)		
Dad w/ College or More	0.263	0.223	0.086	10,073	0.269	0.223	0.115	9,751
	(0.022)	(0.023)			(0.021)	(0.025)		
Mom w/ College or More	0.143	0.126	0.338	10,135	0.145	0.119	0.270	9,688
	(0.017)	(0.018)			(0.016)	(0.020)		
Parents Own Home	0.711	0.732	0.264	10,476	0.716	0.748	0.033	10,202
	(0.015)	(0.016)			(0.013)	(0.013)		
Internet Access at Home	0.916	0.912	0.616	10,272	0.907	0.903	0.580	9,939
	(0.006)	(0.007)			(0.006)	(0.007)		
Standardized Test Score	0.114	0.057	0.181	11,925	-0.100	-0.089	0.827	12,306
	(0.030)	(0.044)			(0.037)	(0.042)		

B. Teacher Characteristics	Female Teachers				Male Teachers			
	with FS	with MS	P-value	Observations	with FS	with MS	P-value	Observations
Teacher Age	36.2	35.8	0.515	15,719	43.0	43.6	0.569	8,406
	(0.575)	(0.567)			(0.963)	(0.811)		
Teacher Experience (year)	11.8	11.4	0.524	15,265	17.0	18.0	0.414	8,315
	(0.605)	(0.614)			(0.975)	(0.952)		
Teachers College Graduate	0.759	0.736	0.467	15,794	0.633	0.614	0.777	8,406
	(0.030)	(0.030)			(0.063)	(0.048)		
Graduate Degree Teacher	0.207	0.204	0.923	15,794	0.454	0.401	0.382	8,406
	(0.029)	(0.032)			(0.052)	(0.057)		
Homeroom Teacher	0.787	0.824	0.216	15,672	0.601	0.660	0.318	8,319
	(0.030)	(0.024)			(0.049)	(0.048)		
Regular Full Time Teacher	0.956	0.962	0.518	15,647	0.967	0.980	0.633	8,360
	(0.013)	(0.011)			(0.022)	(0.015)		

Notes: Each p-value is for a test of equality of means. Standard errors in parentheses are clustered at the school level.

2.3.2 Specifications

To analyze the effect of teacher-student gender interaction, we estimate the following linear regression equation:

$$y_{ijsb} = \beta_0 + \beta_1 fs_i + \beta_2 ft_j + \beta_3 fs_i ft_j + X_{ij}\gamma' + \alpha_s + \alpha_b + \varepsilon_{ijsb}, \qquad (2.1)$$

where y_{ijsb} is the test score of student i who was taught by teacher j in school s for subject b. The test scores are normalized in each subject to have mean zero and variance of one. Because the scores in Korean language, English language and math are pooled together, we also include subject fixed effects α_b. fs_i and ft_j are indicator variables having value of one when student i and teacher j, respectively, are female. X_{ij} is a vector of student and teacher characteristics including indicators for married parents and parental education; teacher characteristics include indicators for graduate degree and graduation from a teachers college, and indicators for teacher experience of two years or below, two to three years, three to four years, four to five years, and five years or more. α_s are school fixed effects, included since random assignment of students is done within schools.

We estimate Equation 2.1 by ordinary least squares (OLS), which produces unbiased estimates given the random assignment of students and teachers to classrooms. Standard errors are clustered at the school level to accommodate correlations among students within the same schools; we obtain similar standard errors clustering at the classroom level or with two-way clustering at the student and teacher level.

β_1 is the average difference in academic achievement for female compared to male students with male teachers, while β_2 indicates the impact of a female versus male teacher on performance for male students. The total effect of having a female teacher for female students can be obtained by adding β_2 to β_3, with β_3 as the differential effect on female students, as compared to male students, of having a female teacher. This last coefficient is the change in the gender gap between female and male students when switching from a male teacher to a female teacher.

2.4 Results

2.4.1 Main Effects

Table 2.3 presents the coefficients from estimating variations of Equation 2.1. We begin in Column 1, with school and subject fixed effects. The coefficient on the female student variable indicates that female students perform better than male student by about 0.15 of a standard deviation on average across Korean language, English language, and math when paired with a male teacher. The change in the performance gender gap between females and males when switching from a male teacher to a female teacher, as indicated by the interaction effect between female student and female teacher, is 0.098 standard deviations. This total effect is comprised of a small and statistically insignificant decrease in male performance of 0.021 standard deviations and an increase in female performance of 0.076. This widening of the gender gap is substantial, representing more than a third of a year of schooling based on the general rule of thumb that 1 percent of a standard deviation of performance is roughly equivalent to 10 days of schooling (Carlsson et al., 2015).

Including teacher background controls in Column 2 does not change the coefficients of our interest much.[14] We replace school and subject effects with include school-by-subject fixed effects beginning with Column 3. In Column 4, we add student fixed effects to test for the presence of unobserved student characteristics correlated with the variables of interest. These also subsume classroom fixed effects and also control for peer effects, since students do not change classrooms. Their inclusion does not change the gender gap appreciably. In Column 5, we follow Fairlie, Hoffmann and Oreopoulos (2014) and in-

[14]Results are similar when including a variety of student background characteristics, but survey nonresponse reduces the sample substantially.

Table 2.3: Main Results

	Single Teacher Only						Single and Multiple Teachers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Student	0.147^{***}	0.144^{***}	0.157^{***}				0.148^{***}
	(0.030)	(0.030)	(0.033)				(0.029)
Female Teacher	-0.021	-0.022	-0.052	-0.025			-0.028
	(0.029)	(0.029)	(0.052)	(0.094)			(0.025)
Female Student × Female Teacher	0.098^{***}	0.105^{***}	0.086^{**}	0.104^{**}	0.093^{*}	0.094^{*}	0.103^{***}
	(0.032)	(0.032)	(0.035)	(0.051)	(0.049)	(0.049)	(0.030)
Constant	-0.093^{***}	-0.110^{***}	-0.047	-0.107	-0.033^{**}	-0.033^{**}	-0.087^{***}
	(0.021)	(0.040)	(0.100)	(0.111)	(0.016)	(0.016)	(0.019)
Observations	24,231	23,580	23,580	23,580	24,231	24,231	30,673
R^2	0.110	0.112	0.123	0.862	0.862	0.859	0.106
School FEs	Yes	Yes					Yes
Subject FEs	Yes	Yes					Yes
Teacher Control		Yes	Yes	Yes			
Sch × Sbj FEs			Yes	Yes	Yes	Yes	
Student FEs				Yes	Yes	Yes	
Cls × Sbj FEs					Yes		
Teacher FEs						Yes	

Notes: Each column represents a separate regression. Columns 1 through 6 are for students taught by a single subject teacher, with each variable as a binary indicator. Column 7 includes students taught by either single or multiple subject teachers, with the female teacher variable representing the fraction of the student's subject teachers who are female. Student and teacher level variables are omitted in Columns 4 through 6 because of collinearity. Standard errors in parentheses are clustered at school level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

clude classroom-by-subject fixed effects to account for the possibility of subject-specific classroom shocks. The results are unchanged. Finally, in Column 6, we add teacher fixed effects to test whether unobserved teacher characteristics are driving our results, despite random assignment. The teacher-student gender interaction coefficient remains the same size and is statistically significant at $p = 0.056$. Taken together with the evidence in Subsection 2.3.1, the stability of this coefficient strongly suggests that the random assignment to classrooms in South Korea is, indeed, in place. As such, the interpretation of our results is free of the potential problems caused by sorting on unobservable characteristics.

Our findings are comparable in magnitude to those in Dee (2007) and Carrell, Page and West (2010). Dee's estimate of the increase in the gap between female and male students when assigned to a female teacher is about 0.092, with opposing positive and negative effects of similar size for female and male students, respectively. While our effect is concentrated on improvements for female students, it is quite similar in magnitude for a similar length of exposure to that year's teachers (about one semester). Carrell, Page and West's effect, for somewhat less than one semester of exposure to a female professor, is 0.097 standard deviations with a reduction in male performance of 0.050 standard deviations.

To show that our results are not affected by the 6,442 observations that were dropped due to students having multiple subject teachers, we include them in Column 5. This specification corresponds to that in Column 1, but the female teacher variable represents the fraction of the student's subject teachers who are female. About 90 percent of these additional observations are groups of two teachers; nearly all of the remaining ones have three teachers. The results are essentially unchanged, with the gender gap increasing by 0.10 standard deviations when all of a female student's teachers are female themselves.[15]

2.4.2 Effects by Subject

The gender gap differs by subject, with female students generally performing substantially better than males in language arts but about even or slightly worse in science and mathematics (OECD, 2015*b*). Teachers' impacts may be greater in mathematics, given negative stereotypes about female mathematical ability; for example, Spencer, Steele and Quinn (1999)'s experimental study shows that negative stereotypes regarding the mathe-

[15]We also estimate the specification in Column (1) excluding 7,964 student-teacher observations at single-sex schools, and an additional 1,620 observations assigned to single-sex classrooms in coeducational schools. The female teacher-female student coefficient for the former sample is 0.087 (s.e. = 0.035) and 0.083 (s.e. = 0.035) for the latter. We also test whether our results differ for students in rural and urban areas and by parental education; no consistent patterns emerge and none of the differences in the interaction variable are statistically significant.

matical ability of female students negatively affects their test scores.

Table 2.4: Results by Subject

	Coefficient	S.E.
A. Main & Interaction Effects		
Female Student	0.336^{***}	(0.040)
Female Teacher	-0.035	(0.039)
Female Student × Female Teacher	0.042	(0.047)
English × Female Student	-0.135^{***}	(0.048)
Math × Female Student	-0.380^{***}	(0.046)
English × Female Teacher	0.049	(0.052)
Math × Female Teacher	0.021	(0.048)
English × Female Student × Female Teacher	0.057	(0.060)
Math × Female Student × Female Teacher	0.042	(0.061)
Constant	-0.156^{***}	(0.027)
B. Change in the Performance Gap		
Korean Language	0.042	(0.047)
English Language	0.099^{**}	(0.045)
Math	0.084^{**}	(0.041)
Observations	24,231	
R^2	0.115	

Notes: Estimates include English and math fixed effects as well as school fixed effects. Standard errors in parentheses are clustered at school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To test whether our results vary by subject, we fully interact the specification in Column 1 of Table 2.3 with indicators for English and mathematics. The coefficients, in Panel A of Table 2.4, show the full set of interactions. We note that female students perform far better than male students in Korean (0.34 standard deviations) and English (0.20 standard deviations), and about evenly in math (-0.04 standard deviations), with the last of these differences being statistically insignificant. In Panel B, we combine the relevant coefficients to calculate the change in the gender gap between female and male performance when switching from a male to female teacher. For Korean language courses, the gender gap between girls and boys does not widen significantly, though it does for English and math; however, there are no statistically significant differences between these effects.

2.5 Evidence on Mechanisms

To investigate the mechanisms underlying the positive impact of female teachers on female students, we examine a series of student responses about classroom interactions, as well as questions about private tutoring asked of parents. There are numerous such questions in the KEDI data, but we chose to focus on those that may distinguish student- and teacher-centered mechanisms. The results in Table 2.5 correspond to the specification in Column 1 of Table 2.3. In Columns 1-4, the dependent variable is an indicator for whether the student agrees or agrees strongly with the following sentiments, respectively: the teacher provides students with equal opportunity to participate in class; the teacher encourages students to express themselves; I feel comfortable asking the teacher a question; and I ask many questions in this class. The first two questions are proxies for teacher-centered mechanisms – that is, they are about the teacher's behavior. The next two questions are about the student's behavior, as are the estimates in Columns 5-7. In Column 5, the dependent variable is a continuous measure of hours of study in that subject (excluding hours spent at tutoring). Column 6, asked of parents, reports the likelihood of receiving tutoring in the subject; note that over 60 percent of students receive tutoring. In Column 7, we examine the effect on the log of tutoring expenditures, conditional on reporting any. This variable, reported by parents as well, provides an indication of tutoring intensity, both in terms of time and personal attention. Finally, Column 8 is the impact on student's self-report that the subject is his or her favorite. This can be influenced by both student- and teacher-centered mechanisms, and provides a useful proxy for the student's overall response to the teacher.

Female students are significantly less likely to feel as if they have an equal opportunity to participate or are encouraged with male teachers, but this negative outcome is eliminated when the teacher is female. On the other hand, while all students report greater comfort in asking questions when the teacher is female, there is no additional effect on female students; they also are somewhat less likely to report asking many questions. There is no effect on hours of study, nor on either tutoring outcome variable. Overall, female students are significantly more likely to report that the subject is their favorite when the teacher is female.[16]

Finally, we examine whether the effects differ by the proportion of female students in the classroom. A greater number of female students means that a female teacher can give less attention to each individual student; on the other hand, a higher proportion of female students may enable the teacher to provide a more welcoming environment for girls. We begin by examining whether the impacts of the teacher-student gender match are greater at single-sex schools. While the interaction effect is somewhat larger (0.037 standard deviations), it is not statistically significant ($p = 0.50$); similar results are obtained when comparing single-sex to coeducational classrooms. Female students in classrooms with above-median numbers of females do perform better with female teachers (0.036 standard deviations), but once again the difference is not significant ($p = 0.57$).

[16]We estimated versions of Table 2.5 with interactions by subject, as in Table 2.4. While effects tended to be larger in math and, to a lesser extent, English classes relative to Korean classes, none of the differences were statistically significant.

Table 2.5: Effects on Student and Teacher Behavior

	Equal Chance to Participate	Teacher Encourages Expression	Comfort Asking Questions	Asks Many Questions	Hours of Study	Receives Tutoring	Log Tutoring Expenditures	Favorite Subject
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Student	-0.050^{***}	-0.056^{***}	-0.029	-0.020^{**}	-0.137	-0.079^{***}	0.031	-0.007
	(0.018)	(0.017)	(0.019)	(0.009)	(0.094)	(0.018)	(0.048)	(0.016)
Female Teacher	0.039^{*}	0.012	0.050^{**}	0.029^{***}	-0.046	-0.014	-0.039	-0.005
	(0.022)	(0.021)	(0.021)	(0.010)	(0.061)	(0.014)	(0.031)	(0.015)
Female Student ×	0.060^{***}	0.063^{***}	0.008	-0.021^{*}	0.063	0.017	-0.080	0.041^{**}
Female Teacher	(0.022)	(0.020)	(0.023)	(0.011)	(0.097)	(0.018)	(0.049)	(0.020)
Constant	0.379^{***}	0.332^{***}	0.314^{***}	0.099^{***}	1.857^{***}	0.623^{***}	2.223^{***}	0.286^{***}
	(0.015)	(0.014)	(0.014)	(0.007)	(0.052)	(0.012)	(0.027)	(0.011)
Observations	23,773	23,737	23,755	24,065	24,227	17,812	6,788	23,900
R^2	0.053	0.055	0.036	0.028	0.044	0.120	0.194	0.028

Notes: Each column represents a separate regression and includes subject and school fixed effects. The response variables for Columns 1 through 4, and 8 are indicators taking value of one if a student agrees or strongly agrees with the statement that the subject teacher gives all students an equal opportunity to participate in class; the subject teacher encourages students to be creative and express themselves; I feel comfortable asking the subject teacher a question when the lecture is difficult to understand; I ask many questions in this class; this subject is one of my favorites. The outcome variable in Column 5 is self-reported study hours per week for the subject, excluding hours spent at tutoring. The outcome in Column 6 is an indicator for receiving tutoring and that for Column 7 is the log of tutoring expenditures. Column 7 is regressed conditional on positive expenditures. Standard errors in parentheses are clustered at school level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We also interact the proportion of female students in a coeducational classroom with the teacher-student gender interaction. Once again, the difference is fairly large but statistically insignificant. For example, a ten percentage point increase in female representation increases the interaction effect by 0.013 standard deviations, with a standard error of 0.023 standard deviations. Without making too much of these differences, they suggest that female teachers may be conducting their classrooms differently in a manner that has a positive impact on female students.

Taken together, the results on female students' responses to female teachers and the somewhat-large effects for classrooms with more female students provide suggestive evidence that the increase in female student performance with female teachers is driven by teacher rather than student behavior.

2.6 Conclusion

Understanding the effect of teacher-student gender interactions on student's academic achievement is important not only for evaluating policies to close the gender gap in academic achievement, but also to enhance understanding of the education production function. However, it is difficult to estimate a student-teacher gender match effect free of selection bias because of the nonrandom sorting of students.

In this study, we estimate the impact of teacher-student gender matches on academic achievement using the random assignment of students in South Korea. We find that the performance gender gap between female and male students increases dramatically when switching from male to female teachers (0.098 standard deviations). Male students do not appear to benefit appreciably from a teacher of the same gender, but female students' performance increases by about 8 percent of a standard deviation when they are taught by a female teacher. This effect is large, and driven primarily by performance in English and mathematics courses. We also provide evidence that teacher behavior drives this increase in student achievement.

Our findings are consistent with the results of Dee (2007) and Carrell, Page and West (2010). Combining these similarities, the random assignment nature of our approach, and the evidence on South Korea's attitudes towards gender equality (Brandt, 2011), we conclude that these interactions reflect genuine changes in the classroom environment that are not necessarily driven by the environment being studied.

3. PERSISTENT EFFECTS OF TEACHER-STUDENT GENDER MATCHES

3.1 Introduction

Over the past forty years, more and more females have moved into many prestigious occupations that were previously dominated by males. Also, women started to exceed men in the attainment of bachelor's degree. However, there has been persistent gender gap in academic outcomes and in employment in science, technology, engineering, and mathematics (STEM) fields. Women's participation in undergraduate computer science and engineering is below 20 percent. Also, females compose only 25 percent of the STEM workforce (Corbett and Hill, 2015).

What is the source of these discrepancies and why do these gaps continue to exist? Many researchers have focused on the role of teacher and student gender. For example, Spencer, Steele and Quinn (1999) shows in their experimental study that negative stereotype on girl's math ability undermines girl's performance on the math tests. Lavy and Sand (2015) find that primary school teachers' gender biases affect the students' academic achievements; using random assignments of teachers and students into classes at Israel primary schools, they show that teachers' biases favoring boys have positive impact on boys' academic performance but negative impact on girls'. Maybe female faculty affects female students deciding the career paths by acting as role models or by making a more favorable environment for them at the university (Rothstein, 1995).

As a policy recommendation to close the gender gap, teacher student gender matches have been discussed frequently among researchers,[17] because of the sizable effects; Lim

[17] Researchers have studied gender matching in various levels of education. At primary school level, Antecol, Eren and Ozbeklik (2015) report negative female teacher effects on female student's math achievement and Muralidharan and Sheth (2016) positive effects on female student's math as well as language course scores. At secondary school level, Dee (2007) and Lim and Meer (forthcoming) find positive effects while Holmlund and Sund (2008) and Cho (2012) report no effects, which we suspect are due to nonrandom sorting. Similarly, in higher education level, Carrell, Page and West (2010) report positive effects

and Meer (forthcoming), using random assignment feature in middle school in South Korea, show switching from a male to a female teacher would increase female student's academic achievement comparing to male student by about 10 percent of a standard deviation, which represents nearly a semester of schooling based on the rule of thumb that a one percent of a standard deviation increase in performance amounts to 10 days of schooling (Carlsson et al., 2015). Dee (2007) uses National Education Longitudinal Study of 1988 to point out that the math gender gap among 13-year-olds would be eliminated by switching from a male to a female teacher.

However, most studies are focused on contemporaneous effects and only a few studies investigate whether the impacts persist in longer term as well. Carrell, Page and West (2010), using random assignment of students to professors in compulsory mathematics and science courses at the United States Airforce Academy (USAFA), estimate the effects of introductory course teacher's gender on longer-term outcomes. They find that higher proportions of female introductory math or science course professors are positively correlated with the student's achievement in the follow-on courses, higher-level math coursetaking, and graduating with a STEM degree, for high achievers. Lavy and Sand (2015) report that primary school teacher's overall stereotypical bias favorable for boys affects the student's high school matriculation exam scores, the likelihood of receiving a matriculation diploma, and the number of successfully completed matriculation exams' units, positively for boys but negatively for girls.

In this paper, we study the longer term effects of teacher-student gender matches at secondary education level, in which the literature lacks knowledge on the longer term effects of teacher-student gender interactions. We avoid nonrandom sorting problem using a unique Korean middle school practice: random assignment of students into a classroom

while Hoffmann and Oreopoulos (2009) find no effects. Another measure to close the gender gap is single sex schooling based on numerous researches such as Hoxby (2000), Whitmore (2005), Lavy and Schlosser (2011), Jackson (2012), Lee et al. (2014), and so on.

each year in which they stay throughout a school day and which subject teachers visit to give them a lesson. Also, panel feature of our data, which tracked seventh graders until their 11th grade, enables us to investigate how the effects of teacher-student gender match change over time.

We find the evidence that switching from a male to a female teacher increases female student's test scores compared to male student, and this gender gap effects on student's academic achievement persist for a long time. Our long lasting gender gap effects are somewhat surprising, since the general rule of thumb in education literature is that teacher effects persist at the rate of 30 to 50 percent, meaning only 30 to 50 percent of the effects remain one year after the exposure to the teacher (Jacob, Lefgren and Sims, 2010). We report the mechanisms behind the persistent effects by showing that female student taught by a female teacher goes to high school of higher quality.

The remainder of the paper is organized as follows: Section 3.2 describes institutional background, data, and identification strategy, Section 3.3 discusses the statistical methodologies, Section 3.4 shows our results, Section 3.5 provides suggestive evidence on mechanisms, and Section 3.6 concludes.

3.2 Data

3.2.1 Institutional Background

Secondary education in South Korea has unique features to exploit in studying Economics of Education in a causal way.[18] First of all, elementary school graduates are entered into a lottery for the assignment of their middle school, which is for seventh through ninth grades, regardless of the middle school being public or private. They can submit their preference list for the middle schools to go within a middle school district, if their super-

[18]Kang (2007) uses random assignment feature of classroom assignment to examine peer effects. Lee et al. (2014) examine the effect of single-sex versus co-educational schooling on academic performance exploiting random assignment of student to middle schools in the Seoul metropolitan area.

intendent allows them to do so (Korea Legislation Research Institute, 2011). However, the students residing in Seoul, with whom our data deal, are not allowed to reveal their preferences, resulting in random assignment of students to school within a middle school district.

Most striking difference in secondary education in South Korea and the United States would be classroom and its assignment. While secondary school students in the United States move to a different classroom for each class, those in South Korea stay in a *physical* homeroom classroom, where each subject teacher rotates to give them a lesson. They are assigned the homeroom classroom at the beginning of academic year, March 1st, and the assignment is done so that each classroom has homogeneous students in terms of academic ability,[19] because of the strong social norms and government policies (Kang, 2007). Lim and Meer (forthcoming) confirm this quasi-random classroom assignment practice by surveying 197 middle schools in their data.

Subject teacher's classroom assignment is done in various ways that are not related with characteristics of teacher or students in the classroom; for example, if there are two subject teachers available for a grade at the school, the one takes odd-numbered classrooms while the other takes even-numbered ones, or the one takes lower-numbered half of them while the other takes upper-numbered half. Lim and Meer (forthcoming) confirm this point too by surveying 197 middle schools in their sample and by showing statistically that teacher assignment is not associated with student's characteristics in the classroom. We also check whether this point holds for our data in Subsection 3.2.3.

The quasi-random student assignment and teacher assignment combined produce the random variation in teacher student matching within a school. Using school fixed effects or school by subject fixed effects, Lim and Meer (forthcoming) show the positive effects

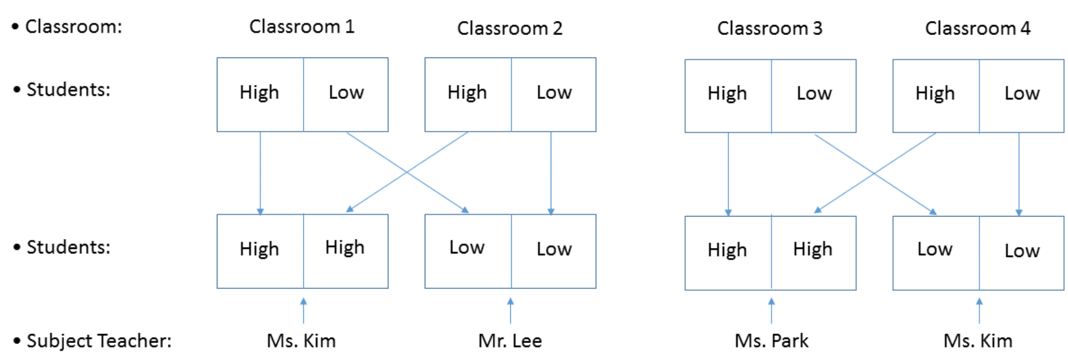
[19] The most common example is to order students by previous year's test scores, and to place the leading student to the first classroom, the second-ranked student to the second classroom, and so on.

of female teacher on female student's academic achievement.

However, over the past decade there has been increasing use of ability tracking in South Korea (Byun and Kim, 2010). In fact, many schools in our data have ability grouping in math and English language, while there are little schools adopting ability grouping in Korean language. The ratios of students belonging to the ability group in math and English language vary by year, with more than 80 percent of students in the sample belonging to ability group for math and English in seventh grade, and around 60 and 30 percent in eighth and ninth grades. If a school adopts ability grouping in a subject, the students move to the classroom of their ability group level for the subject to listen to the lecture, and come back to their original homeroom classroom for the rest of classes for the day; because of the limited classrooms to use in a Korean school, the students are switching their classrooms for the ability grouped class. Schools with ability grouping for a subject, in general, divide students in two classrooms into two or three groups by their ability (i.e. high and low groups or high, middle, and low groups). Figure 1 illustrates this point; students of high ability in Classroom 2 move to Classroom 1 and low ability students in Classroom 1 move to Classroom 2 for the class of their ability.

Recent increase in ability grouping practice threatens the random variation in teacher and student gender. Where there is no ability grouping, we can isolate random variation in teacher student gender matches by conditioning on school by subject fixed effects. However, if we include only school by subject fixed effects with our data where students from different homeroom classrooms but of the same ability group level get together in another place for the class of their ability, the estimated teacher gender effects would be biased; to illustrate with Figure 1, we will end up comparing students taught by Mr. Lee with those taught by Ms. Kim and Ms. Park, if we control for school by subject fixed effects only. That is, we cannot address the student's different academic ability. As is illustrated in Figure 1, most schools have more than two classrooms per grade, and they tend to have two or more classrooms of high or low ability groups in a grade. Accordingly, we can isolate random variation in teacher and student gender by including school by subject by ability group level fixed effects, enabling us to compare students within the same level of ability group.

Figure 3.1: Example of Ability Group Formation in South Korea



Unlike middle school, there are two rounds of admissions to high school in Seoul. Admission to a school in the first round is determined by the school, while admission is given by lottery in the second round (Seoul Metropolitan Office of Education, 2012). In the first round, student can apply for only one of the first round schools, which are 35 magnet high schools,[20] six arts high schools, one athlete high school, and 74 vocational high schools.[21] The school selects students based on their potential ability to learn. If selected, the student cannot apply for the second round high schools. Second round is for those

[20]Magnet high schools in Seoul include three science high schools, one international affairs high school, six foreign language high schools, and 25 autonomous private high schools. Autonomous private high schools are exempt from many education regulations in exchange for not receiving government fund, and attract students of higher ability.

[21]Vocational high schools focus on vocational education, while academic high schools prepare students for the admissions to universities or community colleges. Of course, vocational high school students are not prohibited from applying for the university.

students who failed to gain admissions in the first round or who did not apply in the first round. Second round high schools in Seoul include 19 autonomous public high schools,[22] a few science-focused high schools, a few art-focused high schools, and 183 general academic high schools. In the second round, students are assigned to a school by lottery, based on their choice of a few schools. A student can apply for one autonomous public high school, one science- or one art-focused high school, and four general academic high schools. Applying for an autonomous public high school and a science- or an art-focused high school is optional. Placement for autonomous public high school is determined first and that for general academic high school is determined last. Students living in the same administrative district or high school district[23] as the high school have higher probability to be assigned to the school.

One year after the admissions to the high school, be it academic or vocational, students choose among academic tracks in which students are provided with more focused lesson to a specific field. Most schools provide math-science track and humanities-social science track. Exceptions are that science high school provides math-science track only and foreign language high school provides humanities-social science track only. Humanities-social science track provides more credits of lesson in Korean language and social studies but less credits of lesson in math and science than math-science track. As a result, students in humanities-social science track can learn advanced courses in Korean language and social studies, and math-science track students learn advanced mathematics and science. Because STEM field departments in university requires scores in advanced math and science in College Scholastic Ability Test (CSAT), students planning to study in STEM field in university choose math-science track in high school. Students can freely choose

[22] Autonomous public high schools are also exempt from some regulations like autonomous private high schools. While tuition is lower in the public autonomous high schools, they are subject to more regulation than private autonomous high schools.

[23] In general, a high school district is comprised of two or three administrative districts.

their academic track; past test scores, student gender, and so on are not considered when students are assigned to an academic track. Changing the track after the decision is possible. However, students switching the academic tracks are rare because it would make difficult for them to catch up with their peers who already would have more knowledge on the subjects, leading to lower performance in midterm or final, the results of which are included in the application to university.

3.2.2 Data Set

We use seventh grader panel of Seoul Education Longitudinal Study of 2010 (SELS 2010), which surveyed fourth, seventh, and 10th grade students residing in Seoul in 2010, their parents, teachers of math, English language, and Korean language, principals, and schools. SELS2010 tracks them each year until their 12th grade. We obtained data that tracked seventh graders through their 11th grade.

Students in the seventh grader panel were sampled by stratified two-stage cluster sample design; first, 74 middle schools were randomly chosen from the population of 370 public or private middle schools, excluding two middle schools that are operated by the central government and one athletic middle school in Seoul. Then two classrooms were drawn randomly within the sampled school. Sixty two of the sampled schools are coed, seven are all-boys, and five are all-girls school. 4,658 out of 5,065 target students agreed to participate in the survey and 4,544 students responded to the survey in 2010. The respondents reduced to 4,347, 4,162, 3,541, and 3,394 in 2011 through 2014. The students in the sample advanced to high school in 2013. 3,017 students of them went to academic high school and 524 to vocational high school. 2,893 and 501 students in academic and vocational high schools remained in 2014 (namely, 11th grade).

Subject teachers of math, Korean language, and English language are linked to the students in their seventh through 10th grades. Thus, we use student and subject teacher parings as observations. However, we exploit the variation in teacher-student gender matches in only seventh grade. While ability grouping in seventh grade is determined based on classroom placement test at the beginning of the academic year, ability grouping in eighth and ninth grades are formed by the past academic performance, which are affected by teachers. Also, 10th grade student data do not have ability group level information for which we have to control to obtain random variation in teacher-student matches. Of 13,632 possible teacher student matches, 3,364 observations cannot be linked because of teacher non-response. We drop 74 observations, which have missing values in test scores, student gender, and teacher gender.[24] Among the parings of 10,194, 11 percent are teacher-student parings of male-male, 8 percent are male-female, 44 percent are female-male, and 37 percent are female-female. They represent 4,146 students and 488 teachers. Among them, 1,885 (45.5 percent) students and 402 (82.4 percent) teachers are female.

Student data include standardized test scores in seventh through 11th grades for math, Korean language, and English language; the tests, which covered what students had learned during the first semester of each grade, were administered by Seoul Education Research & Information Institute, and all students in the sample took the same tests at the end of the first semester each year (mid-July). We normalize the test scores to have mean zero and standard deviation of one within each subject and year, for ease of interpretation.

Besides standardized test scores, we examine students' decisions in high school relating to STEM outcomes including math-science track choice and advanced math coursetaking, and aspiration for pursuing STEM major degree. One year after the admissions to the high school, students choose either math-science track or humanities-social science track. Students in different tracks are provided with different courses; while math-science track students can choose to learn advanced mathematics and science, those in humanities-social

[24]We compare remaining 10,194 observations with the dropped 3,438 observations and find no significance differences between them in student's predetermined characteristics (for example, parents' marital status, parent's education, student's height and weight).

science track can choose to learn advanced courses in language arts and social studies. We define advanced math course taking as having taken at least one out of Calculus 2 and Geometry and vector,[25] which are required for those who apply for STEM field in most universities but are not required for those applying for non-STEM field.

3.2.3 Tests of Randomness

In addition to the institutional background justifying the random variation in teacher student gender matches in Korean middle schools, we provide various empirical evidence to validate our identification strategy. First, we examine whether the characteristics of students who are taught by a female versus a male teacher look different. When the students are randomly assigned either to a female or to a male teacher, we expect the observable characteristics of the two groups look similar. We present the sample mean from our data, with each observation as a teacher-student pairing, separately for observations with female and male teachers. Specifically, we regress student characteristics on a female teacher dummy to check whether the difference in any predetermined student characteristics between the two groups are significant at 10 percent level. We cluster standard errors at the school level to accommodate the correlation within a school. Similarly, we check if characteristics of teachers who are linked to female versus male students are different.

In Panel A of Table 3.1, the characteristics for female and male students are presented separately by teacher gender. The student's characteristics are balanced regardless of being taught by a female or a male teacher in seventh grade. Only exception is body mass index based on self-reported height and weight, but the difference is economically not big.

[25]National curriculum revised in 2011, which is applicable to students in our data, lists nine courses in mathematics: Basic math, Math 1, Math 2, Probability and statistics, Calculus 1, Calculus 2, Geometry and vector, Advanced math 1, Advanced math 2. Basic math is for students who lacks in the foundation of math and covers what middle school students learn. Students can learn Probability and statistics, Calculus 1, Calculus 2, or Geometry and vector after finishing Math 1 and Math 2 by the guidance of Seoul Metropolitan Office of Education (Seoul Metropolitan Office of Education, 2011). Advanced math 1 and 2 cover what college students learn, and our data do not have students who took them.

Table 3.1: Comparison of Mean Characteristics

A. Student Characteristics	w/ FT	w/ MT	P-Value	Observations
Female Student	0.46	0.42	0.37	10,258
	(0.02)	(0.05)		
Married Parents	0.87	0.87	0.85	10,029
	(0.01)	(0.01)		
Both Parents Work	0.47	0.46	0.20	9,984
	(0.01)	(0.01)		
Family Income	489.5	480.8	0.65	9,122
	(18.7)	(19.5)		
Parents w/ B.A. or Higher	0.57	0.58	0.67	9,427
	(0.03)	(0.03)		
BMI	19.3	19.6	0.01	9,433
	(0.10)	(0.10)		
Number of Siblings	1.85	1.84	0.47	10,005
(Including Students)	(0.01)	(0.02)		
B. Teacher Characteristics	w/ FS	w/ MS	P-Value	Observations
Female Teacher	0.83	0.80	0.38	10,258
	(0.02)	(0.03)		
Graduate Degree	0.33	0.33	0.90	10,258
	(0.03)	(0.03)		
Teacher's College	0.63	0.62	0.60	10,172
	(0.03)	(0.03)		
Admin Teacher	0.15	0.19	0.03	10,110
	(0.02)	(0.02)		
Over Mid Age	0.53	0.55	0.48	10,170
	(0.03)	(0.03)		
Teaching less than 5 Yrs	0.26	0.26	0.79	9,463
	(0.03)	(0.03)		

Notes: Each p-value is for a test of equality of means. Standard errors in parentheses are clustered at the school level. Unit of Family income is in 10,000 KRW. 10,000 KRW amounts to 865 USD on average in 2010.

Further, we find the difference disappears when controlling for school by subject by ability group level fixed effects, which is justified because the random assignment is done within a school by subject by ability group level cell. In Panel B where we compare teacher characteristics by student's gender, we do not find any predetermined teacher characteristics in seventh grade being significant at 10 percent level.

As another way to check the randomness of the gender variation, we regress subject teacher's gender on student's observable characteristics, controlling for school by subject by ability group fixed effects. The results are shown in Table 3.2. We find that teacher gender in seventh grade is not correlated with student's predetermined characteristics. These results are consistent with random matches between teacher and student.

Table 3.2: Randomness Check: Regression of Teacher Gender

	Coefficient	S.E.
Female Student	0.012	(0.009)
Married Parents	0.003	(0.007)
Family Income	0.000	(0.000)
Dad w/ B.A. or Higher	-0.003	(0.004)
Mom w/ B.A. or Higher	-0.002	(0.004)
Attended Prv. Ele. School	0.011	(0.010)
Number of Siblings	0.004	(0.003)
Observations	$8{,}426$	
R^2	0.798	

Notes: We regress dummy indicating female teacher in seventh grade on student characteristics controlling for school by subject by ability group level fixed effects. Standard errors in parentheses are clustered at school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we show that students are randomly assigned to a physical homeroom classroom, and the subject teacher's characteristics are not associated with the students' ability in the classroom. While we cannot check this point for students in seventh and eighth grades because we do not have classroom information for them, we can do so for students in ninth grade, for whom we have the classroom information.

We exploit resampling techniques,[26] which Carrell and West (2010) use to show students at USAFA are randomly placed into sections within each course / semester with respect to academic ability and professor. The first thing to do is calculating empirical p-value for all classrooms appearing in the sample, for test scores of math, English, and Korean in seventh grade as proxies for a ninth grade student's academic ability. To obtain the p-value for each classroom within each school, we randomly draw, without replacement, 10,000 artificial classrooms of the same size from the sample of all students within the school. For each of the synthetic 10,000 classrooms, we compute total test score of students in it. Then, to each of the real classrooms, we give an empirical p-value, which is calculated as the number of artificial classrooms having total test scores lower than those scores of the observed classroom, divided by 10,000. This is to see whether classroom variation in terms of test scores in seventh grade within a school looks random. We repeat it for test scores of three subjects in seventh grade, obtaining three empirical p-values for each classroom.

We expect the empirical p-values would be uniformly distributed because any p-value will be observed with equal probability under random assignment. To test whether the distribution of the empirical p-values for the three test scores is uniform, we employ a Kolmogorov-Smirnov test and a χ^2 goodness of fit test. We show the results of these tests in Table 3.3A. We reject none of 222 school by subject tests at 5 percent level using the Kolmogorov-Smirnov test and 6 (2.7 percent) using the χ^2 goodness of fit test. Therefore, we do not find evidence of nonrandom assignment of students into classrooms by academic ability.

We also check the random assignment of teachers into classrooms with respect to student's academic ability; for each of three test scores, we regress the empirical p-values on

[26] See Lehmann and Romano (2005) and Good (2006) for details.

Table 3.3: Randomness Check: Test for Institutional Feature

	Math	English Language	Korean Language
A. Test for Student Assignment			
Kolmogorov-Smirnov test (No. failed / total tests)	0/74	0/74	0/74
χ^2 goodness of fit test (No. failed / total tests)	1/74	1/74	4/74
B. Test for Teacher Assignment			
Female teacher	-0.017	-0.048	-0.022
	(0.038)	(0.061)	(0.047)
Graduate school degree	-0.062	-0.020	-0.049
	(0.035)	(0.043)	(0.039)
Full time teacher	-0.002	0.043	-0.062
	(0.048)	(0.056)	(0.062)
Administrative teacher	0.030	-0.029	0.033
	(0.042)	(0.067)	(0.045)
Observations	582	617	598

Notes: Each column represents a separate regression. Dependent variable is the empirical p-value from resampling, for three test scores. Independent variables are teacher gender, dummies for graduate degree, full time teacher, and administrative teacher. Each specification controls for school fixed effects. Standard errors in parentheses are clustered at school level. The Kolmogorov-Smirnov and χ^2 goodness of fit test results indicate the number of tests of the uniformity of the distribution of p-values that failed at the 5 percent level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

average characteristics of teachers visiting the classroom,[27] controlling for school fixed effects to accommodate the random assignment within a school. As is shown in Table 3.3B, none of 24 coefficients is statistically significant at the 5 percent level, suggesting that there is little evidence of nonrandom assignment of teachers into classrooms in terms of student's academic ability. Therefore, we are sure that the institutional feature of random

[27]When a school adopts ability grouping for a subject, one or more teachers will teach students in the same physical homeroom classroom for the subject. For such a case, we average teachers' characteristics for the classroom.

teacher student matches within a school each subject are in place in our data.

3.3 Statistical Methods

To analyze the contemporaneous teacher gender effect, we estimate the following linear regression equation:

$$y_{ijbgs} = \beta_0 + \beta_1 fs_i + \beta_2 ft_j + \beta_3 fs_i ft_j + X_i \delta_1' + T_j \delta_2' + \gamma_{bgs} + \varepsilon_{ijbgs}, \tag{3.1}$$

where y_{ijbgs} is a test score of student i taught by teacher j for subject b in ability group level of g, if any, in seventh grade at school s. The test scores are normalized in each subject to have mean zero and variance of one. Also, we pool together the scores in math, English language, and Korean language to see the average effects. fs_i and ft_j are indicator variables having one if student i and subject teacher j are female, respectively. X_i is a vector of student's predetermined characteristics including dummies for living with both parents, having at least one parent with bachelor's degree or higher, and having both parents being employed. T_j is a vector of teacher characteristics, including teacher's age and dummies for teacher graduating from teacher's school, teacher with master's degree, homeroom teacher, and teacher holding an administrative position at school. We include school by subject by ability group level fixed effects γ_{bgs} to compare students of the same ability in a subject within a school, to ensure that ordinary least squares (OLS) produces unbiased estimates. Standard errors are clustered at the school level to accommodate correlations among students within the same schools.

We focus on the β coefficients. β_1 is the difference in average academic achievement between female and male students when taught by a male teacher. β_2 represents the average difference in performance for male students between being taught by a female teacher and being taught by a male teacher. β_3 indicates the change in the gender gap between female and male students when switching from a male to a female teacher.

We slightly modify Equation 3.1 to examine the effects of teacher-student gender matches in seventh grade on standardized test scores over time:

$$y_{ijbgts} = \beta_0 + \beta_1 fs_i + \beta_2 ft_j + \beta_3 fs_i ft_j + \gamma_{bgs} + \varepsilon_{ijbgts}, \tag{3.2}$$

where y_{ijbgts} is test score in year t = 2, 3, 4, or 5 (namely, eighth through 11th grades), which is normalized within a subject and a year. We note that we control for seventh grade school by subject by seventh grade ability group fixed effects rather than school in year t by subject by ability group in year t fixed effects, to exploit the random variation in seventh grade.

Our primary interest is on β_3 when t = 2, 3, 4, and 5, which represents the gender gap effects in one, two, three, and four years after learning from the teacher, respectively.

While pooling all three subjects in Equations 3.1 and 3.2 as in Dee (2007), we include all three subject teachers' gender to estimate the effects of seventh grade teachers on the longer term outcomes such as academic track choice and advanced course taking in high school.

$$D_{is} = \alpha + \beta_1 fs_i + \sum_{b \in m,e,k} (\beta_2^b ft_j^b + \beta_3^b fs_i ft_j^b) + \gamma_{smek} + \varepsilon_{is}, \tag{3.3}$$

where D_{is} is an indicator variable having value of one when student i chooses STEM related outcomes in 11th grade (e.g. math-science track, advanced math coursetaking, or aspiration to major in STEM field). ft_j^m, ft_j^e, and ft_j^k are dummies indicating female teacher in math, English, and Korean in seventh grade. We include school by math ability group by English language ability group by Korean language ability group in seventh grade fixed effects γ_{smek} to isolate random variation, but expect that too many fixed effects lower the precision of our coefficients of interest.

3.4 Results

3.4.1 Contemporaneous Effects

We first look at the effects of seventh grade teacher on student's standardized test scores in seventh grade in Table 3.4. The coefficients are from estimating variations of Equation 3.1. We begin by most parsimonious model in Column 1. We include school by subject fixed effects in Column 2 and include school by subject by ability group level fixed effects for the rest of the Columns to obtain unbiased estimates. We add controls one by one in Columns 4 through 7 to show the stability of our coefficients of interest.

Table 3.4: Contemporaneous Effects in Seventh Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Student	0.069	0.128^{**}	0.087^{*}	0.095	0.117^{*}		
	(0.061)	(0.061)	(0.045)	(0.057)	(0.063)		
Female Teacher	−0.061	−0.044	0.063	0.054	0.039	0.002	
	(0.078)	(0.121)	(0.077)	(0.084)	(0.093)	(0.114)	
Female Student ×	0.133^{*}	0.145^{**}	0.143^{***}	0.141^{**}	0.132^{*}	0.175^{*}	0.180^{**}
Female Teacher	(0.068)	(0.070)	(0.054)	(0.064)	(0.069)	(0.091)	(0.075)
Observations	10,194	10,194	10,045	8,470	7,792	9,217	10,045
R^2	0.008	0.161	0.370	0.391	0.381	0.827	0.823
Sch × Sbj FEs		Yes					
Sch × Sbj × Abg FEs			Yes	Yes	Yes	Yes	Yes
Student Controls				Yes	Yes		
Teacher Controls					Yes	Yes	
Student FEs						Yes	Yes
Teacher FEs							Yes

Notes: Each column represents a separate regression. Student controls include dummies for both parents living together, both parents working, at least one parent having B.A. degree or higher. Teacher controls include dummies for teacher experience of less than five years, teacher's school graduates, and graduate degree teacher. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

In Column 1, the coefficient on female student dummy indicates that when students were taught by a male teacher, the difference in test scores on average across math, Korean language, and English language between boys and girls are insignificant. In Column 3, insignificant coefficient on female teacher dummy variable means that male student's performance is not affected by the teacher gender. The interaction effect between female student and female teacher, which represents the gender gap change favorable for girls when switching from a male to a female teacher, is 14 percent of a standard deviation. This effect consists of decrease in boy's performance by 0.06 standard deviations, though it is statistically insignificant, and increase in girl's performance by 0.08. In other words, the gender gap effect (β_3) is due to the opposite gender teacher effect for male student ($-\beta_2$) and the same gender teacher effect for female students ($\beta_2 + \beta_3$). The gender gap effect is substantial, considering the evidence that 10 days of schooling increases academic performance by 0.01 standard deviation (Carlsson et al., 2015). The gender gap effects in Column 2, which controls for school by subject fixed effects are almost the same as that in Column 3, suggesting that ability group is formed irrespective of teacher-student gender matches.

In Columns 4 through 7, inclusion of student's and teacher's observable characteristics and fixed effects does not change the gender gap coefficient appreciably, implying that our gender gap effect is not correlated with observed and unobserved student and teacher characteristics. These results reconfirm that the variation in teacher and student gender we use is truly exogenous.

We note that Table 3.4 is the replication of the main result table in Lim and Meer (forthcoming) which use data for ninth grade Korean students in 2004. Using different data set in this paper, we find the magnitude of the effects are basically the same as those in Lim and Meer (forthcoming),[28] implying that the random variation in teacher-student

[28]The gender gap effects are around 0.10 standard deviations across the specifications in Lim and Meer (forthcoming).

gender matching in South Korea is true.

3.4.2 Effects over Time

Next, we examine the effects of female teacher over time as one of the longer-term effects to see how quickly the gender interaction effects fade out as time goes by. Using Equation 3.2, we present the changes over time in seventh grade female teacher's effect on student's test scores in later years in Table 3.5. Columns 1 through 4 correspond to the effects of teacher-student gender matches on standardized test scores in eighth through 11th grades. Each column includes school by subject by ability group level fixed effects in Panel A and student fixed effects and teacher fixed effects as well in Panel B.

Surprisingly, we find that the gender gap effects persist even four years after a student is taught by the teacher. The effects vary slightly over time, but there are no significant differences between the contemporaneous effect and the effects in the following years.[29] The effects last even when we add seventh grade teacher fixed effects and student fixed effects in Panel B. Our finding on the persistent gender gap effects is unusual, because the general rule of thumb in education literature indicates that teacher's impacts on test scores persists at the rate of 30 to 50 percent, meaning only 30 to 50 percent of the effects remain one year later.[30] Meanwhile, the lasting impacts might be one unique characteristic in teacher or professor gender effects. For example, Carrell, Page and West (2010)'s gender gap effect of math and science introductory course professors on follow-on STEM course performance among students with high math ability (0.228 standard deviations) is as good as the contemporaneous effect among them (0.172 standard deviations). Lavy and Sand (2015) also show that the effect of the overall stereotypical bias of a primary school teacher on boy's eighth grade test scores (0.254 standard deviations) is very similar to the effect

[29]P-values are $p = 0.76$, $p = 0.71$, and $p = 0.77$ for the differences between the coefficient on interaction term in Column 3 in Table 3.4 and those in Columns 1 through 4, respectively.

[30]See, for example, Jacob, Lefgren and Sims (2010) and Rothstein (2010).

Table 3.5: Effects over Time

	Dep. Var = Std. Test Scores in			
	8th Grade	9th Grade	10th Grade	11th Grade
	(1)	(2)	(3)	(4)
A. Controls for Sch × Sbj × Grp FEs				
Female Student	0.100	0.180***	0.135*	0.090
	(0.061)	(0.065)	(0.069)	(0.067)
Female Teacher in 7th Grade	−0.066	−0.112*	−0.041	−0.108
	(0.064)	(0.057)	(0.061)	(0.071)
Female Student ×	0.136**	0.103	0.154**	0.155**
Female Teacher in 7th Grade	(0.061)	(0.067)	(0.067)	(0.062)
Observations	9,663	9,206	7,386	7,356
R^2	0.342	0.336	0.283	0.217
B. Controls for Sch × Sbj × Grp FEs, Stu FEs, & Tch FEs				
Female Student ×	0.162**	0.146	0.230***	0.217***
Female Teacher in 7th Grade	(0.069)	(0.093)	(0.078)	(0.079)
Observations	9,663	9,206	7,386	7,356
R^2	0.836	0.852	0.839	0.818

Notes: Each column represents a separate regression. Dependent variables in Columns 1 through 4 are normalized test scores within a subject and a year, in 8th through 11th grades, respectively. Panel A controls for 7th grade school by subject by 7th grade ability group level fixed effects and Panel B adds 7th grade student fixed effects and 7th grade teacher fixed effects on top of them. In Panel B, female student and female teacher dummies are subsumed by student fixed effects and teacher fixed effects, respectively. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

on boy's 12th grade matriculation national exam scores (0.236 standard deviations). In Section 3.5, we discuss the possible mechanisms behind our long lasting effects.

3.4.3 Effects on STEM Outcomes

Although the gender gap in math achievement in secondary education is very small, the gender gap persists in many STEM careers; for example, female participation in undergraduate computer science and engineering is below 20 percent and women compose only 25 percent of the STEM workforce. Given the similar preparedness for STEM major, it would be important to find out the causes for the different preference on STEM field. To test whether teacher gender affects the student's choice on the path to STEM career, we estimate Equation 3.3, using STEM related outcomes in high school as dependent variables.

First off, Column 1 in Table 3.6 presents the results for the effects of seventh grade subject teachers on student's math-science track choice in 11th grade. We do not find significant gender gap effects in any subjects in Column 1, but find significant same gender teacher effect for female students in seventh grade math. This effect suggests that female students are more likely to choose math-science track by 15.1 percent (0.100 + 0.051) when taught by female versus male math teacher. Considering that students who apply for STEM majors in college have to take the tests for subjects taught in math-science track in the college entrance exam, the academic track choice is detrimental in STEM career choice. Thus, the female teacher effect of increasing the probability of choosing math-science track by 15.1 percent could close much of the gender gap in STEM career.

Column 2 reports the effects of seventh grade subject teacher gender on whether student takes advanced math course in the first semester of 11th grade. We define advanced math coursetaking as having taken at least one of the math courses for which students are required to take exam if applying for STEM major but are not required otherwise (namely, Calculus 2 or Geometry and vector). We find that seventh grade math teacher reduce the gender gap between female and male students in the likelihood of taking at least one advanced math course by 9.7 percent. The same gender teacher effect for female student is even bigger; female students are more likely to take advanced math courses by 15.7 percent when they were taught by female math teacher in seventh grade comparing to male teacher. Seventh grade English language teacher has negative gender gap effects in taking advanced math course because they would influence students to take more language courses.

Table 3.6: Effects on STEM Outcomes

	(1) Math Track	(2) Advanced Math	(3) Hope STEM
FS	−0.177	−0.038	−0.124
	(0.108)	(0.106)	(0.158)
Math FT in 7th Grade	0.100	0.060	0.031
	(0.069)	(0.053)	(0.088)
Eng FT in 7th Grade	−0.019	0.196	−0.115
	(0.118)	(0.138)	(0.168)
Kor FT in 7th Grade	0.096	−0.014	0.146
	(0.080)	(0.068)	(0.092)
FS × Math FT in 7th Grade	0.051	0.097*	0.146*
	(0.060)	(0.055)	(0.080)
FS × Eng FT in 7th Grade	−0.008	−0.203**	−0.082
	(0.068)	(0.080)	(0.121)
FS × Kor FT in 7th Grade	−0.068	0.056	−0.116
	(0.084)	(0.074)	(0.096)
Observations	1,635	1,456	1,043
R^2	0.229	0.248	0.280

Notes: Each column represents a separate regression, controlling for school by math ability group by English ability group by Korean ability group in 7th grade fixed effects. Dependent variables are dummies for 11th grade student's choosing math-science track, taking at least one advanced math course, reporting to hope to seek a STEM degree conditional on having decided a major. Advanced math coursetaking is defined as having taken at least one of the math courses until the first semester of 11th grade, for which students are required to take exam if applying for STEM major but are not required otherwise (namely, Calculus 2 or Geometry and vector). Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

Next, Column 3 is for the effects of subject teacher gender on student's aspiration for pursuing STEM major degree, showing even larger gender gap effects in math teacher.

3.5 Evidence on Mechanisms

3.5.1 Is It Mechanical?

We can imagine some scenarios where our gender gap effects on test scores persist mechanically over time. First, female students who taught by a female teacher may stay longer in the sample than those who taught by a male teacher. If test scores are serially correlated, girls who learned from a female teacher in seventh grade are more likely to outperform in later years those taught by a male teacher. Then, our estimates on gender gap effects can remain significant, even though real effects fade out.

We test whether the persistent gender gap effects on test scores is caused by attrition of female students who taught by a male teacher, because our sample size shrinks substantially as time goes by (4,544 students in 2010 to 3,394 in 2014, when our students are in 11th grade). Examining the number of dropped observations by year, around 4.3 percent of students were dropped from the previous year's sample each year in 2011, 2012 and 2014, which is the same rate of students who move out to other cities, go abroad, and are absent from school for a long time. In 2013, 14.9 percent of the students were dropped from the sample of 2012. This seems to be due to the Privacy Act which was effective at the end of 2012. It could be that with the introduction of the new law, students wanted to protect their privacy by not responding to the survey for SELS2010. Also, the law requires students to sign the agreement to provide their personal information, which might have deterred them from responding to the survey. Thus, attrition does not seem to be correlated with our coefficient of interest.

To formally test whether these attrition would be problematic for our estimation of gender gap effects, we regress attrition on student and subject teacher gender in seventh grade. Dependent variables in Table 3.7 are dummies for attrition in eighth through 11th grades. We find that attrition is not correlated with our variables of interest and expect that

Table 3.7: Correlation of Teacher Gender with Attrition

	Dep. Var = Dummy for Attrition in			
	8th Grade (1)	9th Grade (2)	10th Grade (3)	11th Grade (4)
FS	-0.009	0.052	0.035	-0.032
	(0.026)	(0.052)	(0.083)	(0.049)
Pct. FTs in 7th Grade	0.038	-0.041	0.092	-0.077
	(0.036)	(0.045)	(0.120)	(0.070)
FS $\times$ Pct. FTs in 7th Grade	0.014	-0.053	-0.108	0.034
	(0.029)	(0.063)	(0.105)	(0.059)
Constant	0.007	0.075^{**}	0.138	0.103^{*}
	(0.029)	(0.035)	(0.096)	(0.055)
Observations	2,292	2,204	2,113	1,721
R^2	0.114	0.186	0.293	0.196

Notes: Each column represents a separate regression, controlling for school $\times$ math ability group $\times$ English ability group $\times$ Korean ability group in 7th grade fixed effects. The response variables for Columns 1 through 4 are, respectively, dummies for students dropping from the sample in 8th through 11th grades. Percent female teachers is defined the proportion of females in three subject of math, English, and Korean, conditional on there are no missings in the teacher gender. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

we would not suffer from attrition bias.

Next, if female students taught by a female teacher are more likely to be taught by a female teacher in later years as well, the effect of seventh grade teacher will be mixed with the effects of female teacher in later years.

In Table 3.8, we examine whether teacher-student gender interaction in seventh grade affects teacher gender in later grades. Because students are linked to teachers until 10th grade we can show the results for teacher gender only in eighth through 10th grades. Column 3 shows that female students are more likely to be taught by female teacher in 10th grade, reflecting the fact that there are more female teachers in all-girls high school. In none of Columns, the gender gap effects are small and insignificant, showing that our persistent gender gap effects are not driven by the consecutive exposure to female teacher.

Table 3.8: Effects on Future Teacher Gender

	Dep. Var = Dummy for Female Teacher in		
	8th Grade (1)	9th Grade (2)	10th Grade (3)
Female Student	0.026	-0.012	0.198***
	(0.023)	(0.018)	(0.040)
Female Teacher in 7th Grade	0.034	-0.037	0.046
	(0.027)	(0.022)	(0.038)
Female Student × Female Teacher in 7th Grade	-0.021	0.015	-0.041
	(0.024)	(0.019)	(0.039)
Constant	0.806***	0.836***	0.429***
	(0.023)	(0.019)	(0.034)
Observations	9,660	7,603	6,825
R^2	0.475	0.546	0.170

Notes: Each column represents a separate regression controlling for 7th grade school by subject by 7th grade ability group level fixed effects. Dependent variables in Columns 1 through 3 are dummies indicating female teacher in 8th through 10th grades. Standard errors in parentheses are clustered at school level.
* $p < .10$, ** $p < .05$, *** $p < .01$

We also check whether gender matching in seventh grade affects the level of ability group in later grades, in Table 3.9. Dependent variables are dummies for student's belonging to high ability group level(s) in eighth through 10th grades. High ability group level is defined as the ability group level labeled as High, Middle of High, or High of High. Again, we find that the persistent effects are not through the ability group formation in later years.

3.5.2 Changes in Student's Choice

We turn to our attention to student's behavioral change in later years induced by the teacher-student gender interaction in seventh grade. First, female students taught by a female teacher might keep achieving high academic performance by choosing peers or teachers of higher quality. In South Korea, it cannot occur in middle school because of

Table 3.9: Effects on Ability Group in Later Years

	Dep. Var = Dummy for High Level in		
	8th Grade	9th Grade	10th Grade
	(1)	(2)	(3)
Female Student	−0.018	0.049	−0.050
	(0.028)	(0.048)	(0.043)
Female Teacher in 7th Grade	−0.018	−0.086	−0.047
	(0.040)	(0.069)	(0.056)
Female Student ×	0.016	−0.009	0.020
Female Teacher in 7th Grade	(0.029)	(0.051)	(0.048)
Constant	0.495***	0.539***	0.440***
	(0.034)	(0.057)	(0.047)
Observations	3,633	1,711	3,494
R^2	0.450	0.464	0.202

Notes: Each column represents a separate regression, controlling for 7th grade school by subject by 7th grade ability group fixed effects. The response variables for Columns 1 through 3 are dummies indicating a student is in an ability group level which is labeled as a high group in 8th through 10th grades, respectively. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

random assignment of students into classroom and to teachers. However, student has some degree of freedom to choose their own high school. Tables 3.10 and 3.11 examine this point. In Table 3.10 we calculate peer quality by averaging previous years' standardized test scores of students, excluding a student's own score, for each subject teacher.

As is expected we do not find any significant effects on middle school peers in Columns 1 and 2. However, we find large increase in peers' quality. In other words, female students taught by a female rather than a male teacher have peers of higher standardized test scores compared to male students by 15.2 percent of a standard deviation.

In Table 3.11, teacher quality, or teacher value added is calculated by estimating teacher fixed effects in eighth through 10th grades. We obtain it by the following steps:

Table 3.10: Effects on Peer Quality in Later Years

	Dep. Var = Prev. Scores of Peers in		
	8th Grade (1)	9th Grade (2)	10th Grade (3)
Female Student	-0.010 (0.029)	0.013 (0.014)	0.071 (0.047)
Female Teacher in 7th Grade	-0.028 (0.022)	0.015 (0.012)	-0.045 (0.058)
Female Student $\times$ Female Teacher in 7th Grade	-0.002 (0.029)	-0.011 (0.014)	0.152^{***} (0.053)
Observations	9,615	9,028	7,095
R^2	0.656	0.793	0.279

Notes: Each column represents a separate regression, controlling for school by subject by ability group fixed effects. Dependent variables are peer qulity in grades appeared in Column head. Peer quality is defined as average standardized score in previous year, excluding a student herself's or himself's score, for each subject teacher. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

first, for each teacher in gth grade (where g = 8, 9 and 10), we take all the students whom the teacher teaches. For each gth grade student who learns from the teacher, we calculate the leave-out means for $(g-1)$th and gth grade test scores, which is defined as the mean excluding the student's score. Then, we regress the leave-out mean for gth grade on that for $(g-1)$th grade and gth grade teacher fixed effects, controlling for seventh grade school by subject by seventh grade ability group fixed effects.

Columns 1 and 2 in Table 3.11 show that seventh grade teacher-student interaction does not influence teacher quality in eighth and ninth grades within a middle school, reassuring the random matching of students with a teacher. However, gender gap effects on 10th grade teacher value added is large and significant (that is, 0.085 standard deviations). Because the effect of 10th grade teacher value added on test scores in 10th grade is 1.16 standard

Table 3.11: Effects on Teacher Quality in Later Years

	Dep. Var = Teacher Value Added in		
	8th Grade	9th Grade	10th Grade
	(1)	(2)	(3)
Female Student	-0.011	0.009	0.041
	(0.017)	(0.009)	(0.046)
Female Teacher in 7th Grade	0.002	-0.011	-0.029
	(0.020)	(0.016)	(0.038)
Female Student $\times$	0.014	-0.001	0.085^{**}
Female Teacher in 7th Grade	(0.015)	(0.010)	(0.041)
Observations	9,578	8,961	6,766
R^2	0.749	0.719	0.597

Notes: Each column represents a separate regression, controlling for 7th grade school $\times$ subject $\times$ 7th grade ability group. Dependent variables are estimated teacher fixed effects in 8th through 10th grades, which are obtained by the following steps: for each teacher j in gth grade (where g = 8, 9 and 10), take all the students whom the teacher j teaches. Then, for each gth grade student i who learns from the teacher j, calculate the leave-out means for $(g-1)$th and gth grade test scores, which is defined as the mean excluding the student i's score. Next, we regress the leave-out mean for gth grade on that for $(g-1)$th grade and gth grade teacher fixed effects, controlling for 7th grade school $\times$ subject $\times$ 7th grade ability group fixed effects. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

deviations in our data, this higher teacher quality explains about 65 percent ($= \frac{0.085\times1.16}{0.155}$) of the effect of seventh grade gender matching on 10th grade test scores.

Because the change in peers and teacher quality is driven by student's high school choice, we examine the high school output quality in Table 3.12. Dependent variable in Column 1 is percent 11th grade students who achieved Above Basic Performance in National Assessment of Educational Achievement (NAEA) test, which was taken by 11th graders in the school when students in our sample were 10th grade. A student is categorized as Basic Performance when he or she understands 20 to 50 percent of what is expected to be achieved. Dependent variable in Column 2 is percent of 11th grade students with Below Basic Performance in NAEA test, and that in Column3 is percent of 12th graders who were admitted to university or junior college for 2013 academic year, when students in our sample was 10th grade. Columns 1 and 2 are from pooled regression for all three subjects.

Table 3.12: Effects on High School Quality

	(1) Above Basic	(2) Below Basic	(3) College Goers
Female Student	0.005	-0.002	0.027
	(0.008)	(0.003)	(0.033)
Female Teacher in 7th Grade	-0.015^{*}	0.007^{***}	
	(0.009)	(0.002)	
Female Student $\times$ Female Teacher in 7th Grade	0.035^{***} (0.011)	-0.010^{***} (0.004)	
Pct. Female Teachers in 7th Grade			0.057
			(0.073)
Female Student $\times$ Pct. Female Teacher in 7th Grade			0.005 (0.041)
Constant	0.799^{***}	0.048^{***}	0.527^{***}
	(0.007)	(0.002)	(0.057)
Observations	6,576	6,580	1,499
R^2	0.265	0.296	0.374
Sch^{7th} $\times$ Subj $\times$ Grp^{7th} FEs	Yes	Yes	
Sch^{7th} $\times$ Eng Grp^{7th} $\times$ Math Grp^{7th} $\times$ Kor Grp^{7th} FEs			Yes

Notes: Each column represents a separate regression. Dependent variable in Column 1 is percent of 11th grade students with Above Basic Performance in NAEA test, that in Column2 is percent of 11th grade students with Below Basic Performance in NAEA test, and that in Column3 is percent of 12th grade students who were admitted to university or junior college for 2013 academic year, when students in our sample was 10th grade. We note that Columns 1 and 2 are from pooled regression for all three subjects. The NAEA test was taken by 11th graders in the school when students in our sample were 10th grade. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

Constant term in Column 1 indicates that when male students were taught by a male subject teacher in seventh grade, 80 percent of students are categorized as Above Basic

Performance in the subject in their high school. While the seventh grade teacher-student gender matching has mild effect on percent students with Above Basic Performance, the effect is substantial for Below Basic Performance. In other words, female students taught by a female versus a male teacher in seventh grade go to high school with lower percent students of Below Basic Performance by 1 percent points compared to male students, when the mean is 4.8 percent for male students taught by a male teacher in seventh grade. We do not find any significant effect on percent college goers in the high school, because high achieving students tend to repeat the college application to get into better university.

3.6 Conclusion

Researchers as well as policy makers have tried to find out the nature of and the ways to close the gender gap, especially in STEM fields. Recently, growing attention has been given on teacher-student gender matches. In terms of policy, the importance of knowledge on the longer term as well as contemporaneous effects is huge. However, data availability has prevented researchers from delving into this matter.

With the help of SELS2010, which provides conditions similar to an experiment free of nonrandom sorting and attrition bias, we are able to document causal relationship between teacher-student gender interaction and student's academic achievement in the longer run. In this paper, we find that the gender gap effect of contemporaneous teacher on test scores in seventh grade (0.143 standard deviations) is very close to that in ninth grade (0.098 standard deviations) in Lim and Meer (forthcoming), which use different data set from SELS2010. The effect size is also in line with Dee (2007)'s effect for eighth graders (0.092 standard deviations) and Carrell, Page and West (2010)'s effect for university students (0.097 standard deviations). Also, we find those effects persist even four years after the exposure to the teacher and provide evidence on the possible mechanisms behind the persistence. Namely, our lasting effects are not driven by attrition of students who were not taught by a female teacher, consecutive matching to female teachers, or consecutive status in higher ability group. Instead, our persistent effects are due to the student's attitude on learning and high school choice. Importantly, we find teacher-student gender matches influence student's academic choice on the path to STEM fields. These findings shed light on the importance of teacher-student gender matches in future policies to close the gender gap, especially in STEM fields.

4. HOW DO PEERS INFLUENCE BMI? EVIDENCE FROM RANDOMLY ASSIGNED CLASSROOM PEERS IN SOUTH KOREA

4.1 Introduction

The proportion of overweight 15-year-old children in OECD countries has steadily risen since 2000 (OECD, 2015*b*). Also, over the past 30 years, childhood obesity has more than doubled in the United States (Ogden et al., 2014). Overweight children are more likely to be obese as adults, and medical spending attributed to obesity is immense. Finkelstein, Fiebelkorn and Wang (2003) suggest that the cost was nearly 80 billion dollars, or 9 percent of the U.S. medical expenditures in 1998.

The fact that obesity has increased at all income levels in the United States (Chang and Lauderdale, 2005) highlights the importance of social factors rather than individual characteristics for the explanation. Because adolescents are heavily influenced by their peers, peer effects may play a role.

However, empirical research on peer effects is difficult because of well-known issues such as self-selection, common environmental factors, and reflection problems.[31] Christakis and Fowler (2007) use 32 years of measured height and weight data in Framingham Heart Study to show that own chance of becoming obese increases by 57 percent if his or her friend is obese. They try to control for selection by including own lagged obesity status. Using the fact that their effect is not affected by the distance between own and friend's home, they argue their effect is not due to the common environment factors. However, there were some arguments about the validity of their identification strategy, especially about whether they properly address the common environment factors.[32] Also,

[31] See Manski (1993) and Epple and Romano (2011) for the details.

[32] See Cohen-Cole and Fletcher (2008*a*), Fowler and Christakis (2008), and Cohen-Cole and Fletcher (2008*b*) for details on their debate.

they do not deal with reflection problem.

After Christakis and Fowler (2007), there have been two major strategies to address the three empirical challenges. First is to use instrumental variable. Trogdon, Nonnemaker and Pais (2008) use peers' birth weight and peers' parents' obesity status to instrument for the peers' weight and Mora and Gil (2013) instrument for friends' BMI with the characteristics of the respondents' friends-of-friends who are not friends with the respondent. However, IV still can be correlated with unobserved peer variables that enter into the peer selection process. Another way is to use institutional setting where peers are randomly assigned. Carrell, Hoekstra and West (2011) use the fact that students are randomly assigned to squadron in United States Air Force Academy and use peer's high school fitness score to show the peer effects on college fitness score. Yakusheva, Kapinos and Eisenberg (2014) leverage data from two anonymous universities where students are randomly assigned to dormitory room. Using baseline weight, they find peer effects on female college student's weight gain.

In this paper, we avoid nonrandom sorting problem by using a unique Korean middle school practice: random assignment of students into a physical homeroom classroom where they stay to take courses with the same classmates for a day throughout a school year and where subject teachers visit to give them a lesson.[33] Because Korean middle school students spend all day with their classroom peers, the classroom peers are an appropriate social network to be examined. To address reflection problem and common environmental factors, we use peers' number of siblings as an instrumental variable for peers' average BMI. A number of studies (e.g. Hesketh et al., 2007; Chen and Escarce, 2010; Haugaard et al., 2013; and de Oliveira Meller et al., 2015) find the number of siblings is highly correlated with child's BMI and the likelihood of being obese. Possible mechanisms of the

[33] Carrell, Hoekstra and West (2011), Yakusheva, Kapinos and Weiss (2011), and Yakusheva, Kapinos and Eisenberg (2014) are similar to our study in that they avoid the selection bias by using natural experiment.

high correlation between the number of siblings and a child's BMI can be found in our data in which a child with siblings tend to have higher level of physical activity and lower food intakes. Also, the fact that peers' number of siblings cannot directly affect a student's own health condition shows that peers' number of siblings is not correlated with student's characteristics.

Our instrumental variable estimate indicates that a one unit increase in peers' BMI increases a student's BMI by 0.83 units. Also, the effect of seventh grade peers' BMI on student's BMI in eighth grade is still significant and positive (0.51). The contemporaneous and longer term effects of seventh grade peers' BMI are stable when we include the student's, peers', and teacher's characteristics.

While adult's overweight rates in South Korea are among the lowest in OECD countries, overweight men has increased rapidly for a decade (OECD, 2010).[34] Furthermore, ratio of overweight boys in 2013 exceeds average ratio for boys in 33 OECD countries (OECD, 2015*b*).[35] Thus, our results will be relevant for other developed countries.

The remainder of the paper is organized as follows: Section 4.2 describes our data and identification strategy, Section 4.3 discuss the statistical methodologies, Section 4.4 shows our results, Section 4.5 is for possible mechanisms, and Section 4.6 concludes.

4.2 Data

4.2.1 Data Set

We use seventh grade data of Gyeonggi Education Panel Study (GEPS2012), which surveyed 4,051 seventh grade students in middle schools in Gyeonggi province that surrounds Seoul, South Korea. Students were sampled by two-stage cluster sampling design;

[34] 36 and 27 percent of males and females were overweight in 2008 in South Korea with OECD averages being 57 and 46 percent for male and female. In 1998, those ratios for Korean males and females were 25 and 26 percent.

[35] In 2013, overweight boys and girls are 26.4 and 14.1 percent in South Korea, while OECD averages are 24.3 and 22.1 percent for boys and girls, respectively.

first, 63 schools were chosen from the population of 624 middle schools in Gyeonggi province. Then, two classrooms were drawn within each school, and all students in the classrooms were surveyed. GEPS also surveyed parents, homeroom teachers, principals, and schools. Each student is linked to their homeroom teacher.[36]

Each year students were asked to report their height and weight, with which we construct BMI for each student. Because height and weight are self-reported, BMI would be calculated with error. However, the mean measurement error would be zero, because students would want to report their height and weight so that their calculated BMI approaches mean BMI. Using peers' number of siblings as an IV for peers' BMI, we circumvent the measurement error problem.

BMI has been criticized as sometimes misclassifying an individual as obese or overweight when he or she is muscular because it cannot distinguish adipose tissue from muscle, bone, and other lean body mass (Burkhauser and Cawley, 2008). Nevertheless, BMI is still useful in defining adolescent's overweight because it is convenient and can give consistent measure with adult's overweight, which is defined with BMI (Bellizzi and Dietz, 1999).

The data also include the information on the student's diet and exercise routine (e.g. hours for exercise per week, the number of eating breakfast per week, and the number of having dinner with family per week), which allows us the opportunity to study the possible mechanisms behind the peer effects.

Starting with 4,051 students, we drop 128 observations that have missing height or weight information. We have final sample of 3,909 students after dropping 14 additional students who do not have classroom information or whose parents report that the sum of male and female children at home is zero. We find the student and homeroom teacher

[36]Since a homeroom teacher is responsible for managing the classroom to which the students belong, homeroom teachers are much more important than subject teachers in terms of BMI.

characteristics do not look different between the kept and dropped students in the sample.

Table 4.1: Summary Statistics

Variable	Mean	(Std. Dev.)	Min	Max	N
BMI in 7th Grade	19.4	(3.0)	11.8	34.1	3,909
8th Grade	19.9	(3.1)	10.5	39.5	3,556
9th Grade	20.2	(3.1)	12.6	35.9	3,467
Number of Siblings	2.15	(0.73)	1	18	3,909
Female Student	0.48	(0.50)	0	1	3,909
Peers' BMI in 7th Grade	19.4	(0.7)	17.1	21.8	3,909
8th Grade	19.9	(0.7)	18.2	22.8	3,556
9th Grade	20.2	(0.7)	18.6	22.5	3,467
Peers' Number of Siblings	2.14	(0.15)	1.74	2.81	3,909

4.2.2 Randomness Check

First, we conduct a series of Pearson's χ^2 tests for the independence of various students' characteristics and their assigned classroom for each of seventh through ninth grades. Tested student characteristics include student's gender, number of siblings, father's and mother's education, parents' marital status, as well as whether parents own their own home. The maximum number of siblings is 18, parents' education has six categories, and the rest of the characteristics are dummy variables. 11 of 378 tests are not available in seventh grade because five schools are single-sex (5 tests) and one school has only one classroom sampled (1×6 tests). Since students in the school with only one classroom sampled in seventh grade were distributed to multiple classrooms in eighth and ninth grades, only five tests are unavailable in eighth and ninth grades.

Table 4.2: Number of Rejections in Pearson's χ^2 Tests

	7th Grade (367 Tests)			8th Grade (372 Tests)			9th Grade (372 Tests)		
Significance Level	1%	5%	10%	1%	5%	10%	1%	5%	10%
Female Student	0	0	0	0	0	1	0	1	3
Number of Siblings	1	1	2	0	2	4	0	1	3
Dad Education	0	0	1	0	2	3	2	4	7
Mom Education	0	0	2	1	5	8	1	2	5
Married Parents	0	2	7	1	3	4	0	4	6
Own Housing	1	3	4	0	2	6	1	2	8
Sum	2	6	16	2	14	26	4	14	32
Percent Rejected	0.5	1.6	4.4	0.5	3.8	7.0	1.1	3.8	8.6

Table 4.2 shows the number of rejections for the null hypothesis of independence in a series of Pearson's χ^2 tests at 1, 5, and 10 percent significance level. For each significance level, the rejection rates are below the significance level. Hence, we conclude there is little evidence of nonrandom assignment of students into classroom with respect to student's observable characteristics.

We also test whether mean characteristics of classroom peers' are correlated with own characteristics. As Guryan, Kroft and Notowidigdo (2009) point out, in general, the test for random peer assignment by regressing own characteristic on peers' characteristic is not well behaved because of the negative bias that is inherent in the test. Intuitively, when peers are randomly assigned, the classroom averages for student characteristics will be balanced across the classrooms. If, for example, own family income is higher than classroom mean, classroom peers' family income, which is defined as classroom average for family income excluding own one, will be lower comparing to that for a student with lower family income than classroom mean. It suggests negative correlation between own and peers' family income in spite of the random assignment. Guryan, Kroft and Notowidigdo (2009) show, using Monte Carlo study, the over-rejections that the regression of own characteristic on peers' characteristic rejects the independence at 5 percent level more than 33 percent of the time, and show the over-rejection problem is solved with the inclusion of leave-out mean of population from which peers are drawn (that is, school average family income excluding own one in our study).

Following Guryan, Kroft and Notowidigdo (2009), we regress the following equation to test the independence of peers' and own characteristics:

$$y_{ics} = \alpha + \beta \frac{1}{n_c - 1} \sum_{j \neq i, j \in c(i)} y_{jcs} + \gamma \frac{1}{n_s - 1} \sum_{k \neq i, k \in s(i)} y_{js} + \lambda_s + \varepsilon_{ics}, \tag{4.1}$$

where y_{ics} is own characteristic in classroom c in school s. n_c and n_s are the numbers of students in classroom and school, respectively. λ_s is school fixed effects to accommodate the random assignment within a school.

Table 4.3: Regression of Own Characterisitcs on Peer Characteristics

	(1) FS	(2) # of Siblings	(3) Parents BA +	(4) Married Parents	(5) Owning Housing	(6) Familiy Income
Avg. Class Peers Characteristic	-0.115	-0.021	-0.041	-0.046	0.034	-0.016
	(0.138)	(0.049)	(0.032)	(0.061)	(0.031)	(0.019)
Observations	4,046	4,046	3,803	4,006	3,973	3,965
R^2	0.955	0.957	0.956	0.941	0.951	0.951

Notes: Each cell represents a separate regression, including school fixed effects and leave-out mean of school peers' characteristic, which is defined as school mean for the relevant characteristic excluding own characteristic. We include the school peers' leave-out mean to correct the bias that is inherent in typical tests for random assignment of peers, following Guryan, Kroft and Notowidigdo (2009). Dependent variables in Columns 1 through 6 are dummy for female student, number of siblings, dummy for both parents having B.A. degree or higher, that for parents being married, that for parents owning own housing, and family income. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.3 shows the results with dependent variables in Columns 1 through 6 being dummy for female student, number of siblings, dummy for both parents having bachelor's degree or higher, that for parents being married, that for parents owning own housing, and family income, respectively. We find none of the characteristics are correlated with classroom peers' characteristics.

If a homeroom teacher with specific characteristics are more likely to be assigned to the students with higher propensity to gain weight, our peer effects estimates will be confounded with effects from such teachers. To check whether homeroom teacher assignment is done irrespective of student and teacher characteristics, we regress average classroom characteristics on each of teacher characteristics. We include school fixed effects to accommodate the random assignment within a school. Because the number of homeroom teachers matched with students in seventh grade in our data is as small as 120, we regress on each teacher characteristics one by one. The unit of family income in Column 5 is one million KRW, which is equivalent to 930 USD at the end of 2012.

Table 4.4 shows only one out of 25 regressions indicates significant correlation between teacher and classroom characteristics at 10 percent level.

4.3 Specifications

The basic regression equation to estimate the effects of peers' BMI is:

$$BMI_{ics}^{7th} = \alpha + \frac{1}{n_c} \sum_{j \neq i, j \in c(i)} (\beta BMI_{jcs}^{7th} + X_{jcs}\gamma') + X_{ics}\delta' + HT_{cs}\theta' + \lambda_s + \varepsilon_{ics}, \qquad (4.2)$$

where BMI_{ics}^{7th} is the BMI for seventh grade student i, who is assigned to classroom c at school s. $\frac{1}{n_c} \sum_{j \neq i, j \in c(i)} BMI_{jcs}^{7th}$ is the average BMI of seventh grade student i's classroom peers. X_{ics} and HT_{cs} are vectors of student and homeroom teacher characteristics. Student characteristics include student gender, family income, whether both parents live together,

Table 4.4: Regression of Classroom Characteristics on Teacher Characteristics

	(1) Number of Siblings	(2) Father BA +	(3) Mother BA +	(4) Both Parents	(5) Familiy Income
Female Teacher	−0.037	0.010	0.016	−0.005	−0.432
	(0.047)	(0.022)	(0.024)	(0.023)	(0.286)
Teacher Age over 40	0.061	0.017	0.028	−0.022	−0.130
	(0.043)	(0.021)	(0.026)	(0.037)	(0.402)
Teacher Experience Less than 5 Years	−0.033	−0.018	0.006	0.005	0.160
	(0.039)	(0.020)	(0.028)	(0.027)	(0.369)
Post Graduate Degree	0.032	0.005	0.021	−0.030	0.059
	(0.030)	(0.020)	(0.017)	(0.019)	(0.143)
Teacher's College	0.062*	−0.009	0.011	−0.015	−0.132
	(0.035)	(0.018)	(0.020)	(0.022)	(0.218)

Notes: Each cell represents a separate regression, controlling for school fixed effects. We regress each of student characteristics appear in the Column heading on each of teacher characteristics. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

and whether parents have bachelor's degrees or higher. $\frac{1}{n_c} \sum_{j \neq i, j \in c(i)} X_{jcs}$ is the peers' average characteristics, which are leave-out means of student characteristics. Homeroom teacher characteristics include teacher gender and dummies for teacher experience being less than five years, teacher having graduate degree, and teacher graduating from teacher's school. λ_s are school fixed effects to account for the random assignment of students into classroom within a school. ε_{ics} is the error term.

β will capture the endogenous peer effects in Manski (1993)'s framework. However, the β coefficient estimated from Equation 4.2 will be biased because of the reflection problem and common environmental factors. That is, because a student and peers affect each other's BMI simultaneously and they share the same proximity to cafeteria or playground, the peer effects will be confounded with those effects. To address this problem, we instrument peers' BMI with peers' number of siblings; a number of studies (e.g. Hesketh et al., 2007, Chen and Escarce, 2010, Haugaard et al., 2013, and de Oliveira Meller et al., 2015) find that the number of siblings is highly correlated with child's BMI and the likelihood of being obese. The high correlation between the number of siblings and a child's BMI can be due to the facts that a child with siblings tend to have higher level of physical activity and lower food intakes. These facts were shown by the data in a few studies (e.g. Jacoby et al., 1975 and Bagley, Salmon and Crawford, 2006) as well as our data. Hesketh et al. (2007) explains that parenting practices may differ for families with and without siblings; single-child families may have more restriction on outdoor play and fewer restriction on food. Figure 4.1, which depicts the average BMI by the number of siblings in our data, shows the clear negative relationship, when excluding outliers with only one observation. Furthermore, peers' average number of siblings is arguably exogenous to a student's own health condition.

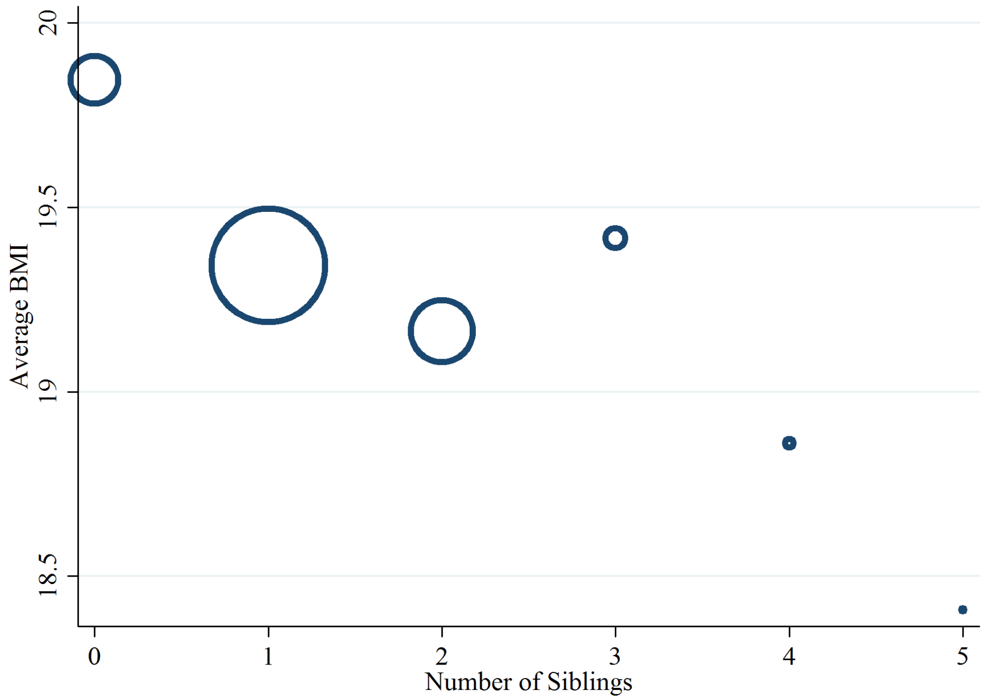
If the classroom with bigger number of siblings has different student or teacher characteristics from that with smaller number of siblings, peer effects estimates might be due to the different peer characteristics, and the instrumental variable approach will be harmed. To see if this is the case in our data, we regress student and teacher characteristics on peers' number of siblings. Table 4.5 shows only one characteristic is significant at 10 percent level, and the effect is economically small; when we regress peers' number of siblings on family income, we find the coefficient on family income is 0.001, meaning increase in family income by one million KRW, 930 USD at the end of 2012, would increase peers' number of siblings by 0.001.

Thus, we use the following as first stage equation:

$$\frac{1}{n_c}\sum_{j\neq i, j\in c(i)} BMI^{7th}_{jcs} = \gamma + \delta\frac{1}{n_c}\sum_{j\neq i, j\in c(i)} S_{jcs} + \theta S_{ics} + \eta_s + \epsilon_{ics}, \qquad (4.3)$$

where S_{ics} and S_{jcs} indicate number of siblings for student i and $j \neq i$.

Figure 4.1: Average BMI by Number of Siblings



Notes: Marker size is proportional to the number of observations. We drop in the plot four points with only one observation.

4.4 Results

4.4.1 Contemporaneous Peer Effects

Table 4.6 presents the coefficients from estimating variations of Equation 4.2 with Column 1 reporting the naïve ordinary least squares (OLS) result, the rest of the columns two-stage least squares (2SLS). We include school fixed effects in each column because random assignment is done within a school.

In Column 1, the OLS coefficient is negative, which reflects the students' BMIs are balanced across classrooms within a school. If student characteristics are perfectly balanced across classrooms and the number of students is the same for all classrooms, we

Table 4.5: Regressions of Characteristics on Peer Number of Siblings

A. Own Student Characteristics	Female Student (1)	Dad BA + (2)	Mom BA + (3)	Both Parents (4)	Own Home (5)	Family Income (6)
Peers' # of Siblings	-0.043	-0.038	-0.000	-0.057	0.003	-1.776^{*}
	(0.036)	(0.057)	(0.061)	(0.065)	(0.083)	(0.910)
Constant	0.579^{***}	0.515^{***}	0.302^{**}	0.966^{***}	0.622^{***}	8.573^{***}
	(0.076)	(0.123)	(0.130)	(0.140)	(0.178)	(1.950)
Observations	4,037	3,924	3,897	4,036	3,969	3,962

B. Homeroom Teacher Characteristics	Female Teacher (1)	Teacher over 40 (2)	Low Experience (3)	Teacher's College (4)	Post Graduate (5)	Admin Teacher (6)
Peers' # of Siblings	-0.189	0.356	-0.205	0.720	0.485	-0.042
	(0.399)	(0.285)	(0.286)	(0.468)	(0.375)	(0.258)
Constant	1.219	-0.476	0.645	-0.855	-0.651	0.153
	(0.853)	(0.610)	(0.612)	(1.002)	(0.802)	(0.554)
Observations	3,888	3,888	3,888	3,851	3,888	3,747

Notes: Each column represents a separate regression, controlling for school fixed effects. In Panel A, dependent variables are dummies for female student, dad with bachelor's degree or higher, mom with bachelor's degree or higher, both parents living together, parents' owning home, and family income of which unit is one million KRW, which is equivalent to 930 USD at the end of 2012. In Panel B, dependent variables are dummies for female teacher, teacher older than 40, teacher with less than 5 years of experience, teacher graduated from a teacher's college, post graduate teacher, and teacher with an administrative position. Standard errors in parentheses are clustered at school level.

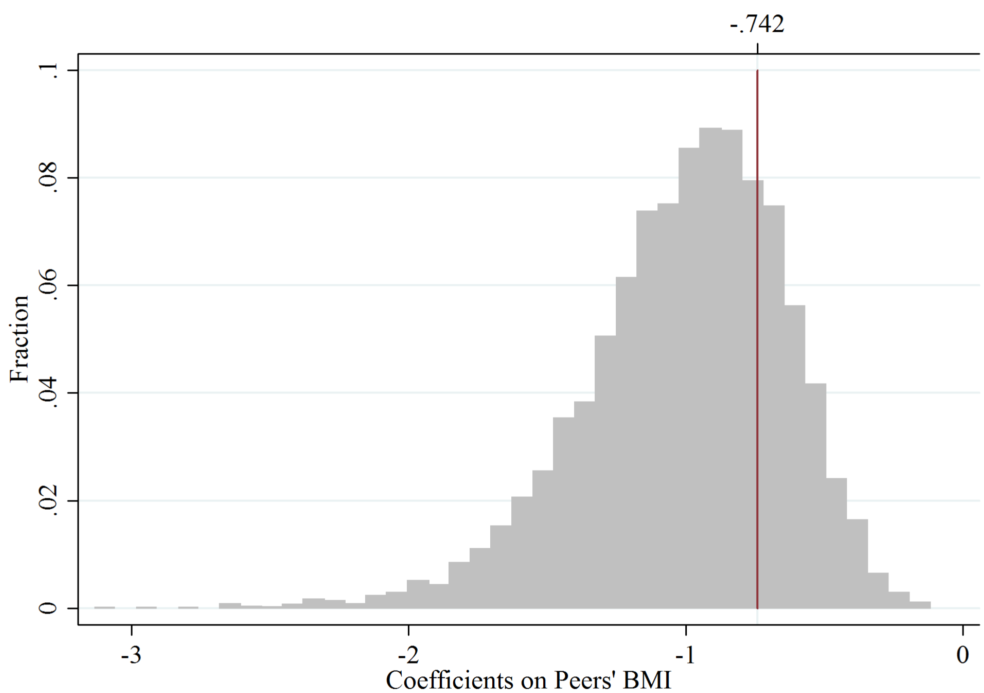
* $p < .10$, ** $p < .05$, *** $p < .01$

mechanically have negative coefficient. A simple algebra shows this fact. Suppose the sum of students' BMIs and the number of students in each classroom are $\bar{y}$ and n and the same across classrooms. We define average classroom peers' BMI for student i as $X_i = \frac{\bar{y}-y_i}{n-1}$, where y_i is own BMI. Then, regression of own BMI on peers' BMI is the same as estimating the following equation:

$$
\begin{aligned}
y_i &= \alpha + \beta X_i \\
&= \alpha + \beta(\frac{\bar{y} - y_i}{n-1}) \\
&= \alpha + \beta\frac{\bar{y}}{n-1} - \beta\frac{1}{n-1}y_i.
\end{aligned}
\tag{4.4}
$$

Because $-\beta\frac{1}{n-1}$ is always 1, we have $\beta = -(n-1) < 0$.

Figure 4.2: Distribution of Coefficients under Random Assignment



We also check whether the negative OLS coefficient in Column 1 looks more extreme than synthetic coefficients made of randomly resampled students from the same school. First, we take students within a school and randomly reassign them into the classrooms of the same size as the existing classrooms. Using the artificial classrooms, we regress own BMI on peers' BMI controlling for school fixed effects, and save the coefficient on peers' BMI. After repeating this procedure 10,000 times, we obtain the sampling distribution of the coefficients under the null hypothesis that peers' BMI is random (Figure 4.2). There are 2,470 coefficients in the distribution that are larger than or equal to -0.742, and we conclude the observed coefficient is due to random assignment rather than chance.

Table 4.6: Endogenous Peer Effects on BMI

	OLS	2SLS			
	(1)	(2)	(3)	(4)	(5)
A. OLS & 2SLS Results					
Avg. Peers' BMI	-0.742^{***}	0.831^{***}	0.761^{***}	0.753^{***}	0.782^{***}
	(0.262)	(0.095)	(0.199)	(0.165)	(0.177)
Own # of Siblings		-0.214^{***}	-0.154^{**}	-0.154^{**}	-0.171^{**}
		(0.075)	(0.073)	(0.074)	(0.077)
Observations	3,909	3,909	3,630	3,630	3,464
R^2	0.052	0.008	0.019	0.019	0.018
B. 1st Stage Results					
Peers' # of Siblings		-1.315^{***}	-1.291^{***}	-1.398^{***}	-1.451^{***}
		(0.480)	(0.477)	(0.446)	(0.504)
Own # of Siblings		-0.033^{**}	-0.032^{*}	-0.037^{**}	-0.037^{**}
		(0.017)	(0.017)	(0.016)	(0.018)
1st Stage F-stat		7.52	7.32	9.82	8.28
School FEs	Yes	Yes	Yes	Yes	Yes
Student Ctrl			Yes	Yes	Yes
Peer Ctrl				Yes	Yes
Teacher Ctrl					Yes

Notes: Each column represents a separate regression. Dependent variable is a student's BMI and and independent variable is mean peers' BMI, which is insturmented with average peers' number of siblings in the 2SLS model. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

In Column 2, we find that a one unit increase in peers' average BMI increases a student's BMI by 0.83 units with strong correlation of the instrument and peers' BMI in the first stage (Column 2 in Panel B). Though first stage F-statistic is slightly lower than 10, the rule of thumb number to check for weak IV, we do not expect we would suffer from it; an estimate with Weak IV converges to OLS estimate, but our estimate in Column 2 is clearly different from that in Column 1. The effect is larger than Trogdon, Nonnemaker and Pais (2008)'s and Mora and Gil (2013)'s effects, which are 0.52 and 0.63, respectively. The effect is also comparable to Christakis and Fowler (2007)'s finding that if a friend is obese, the own chances of being obese increases by 57 percent. Our larger effects might be due to the fact that classroom peers spend most of the time together in the same classroom, considering Carrell, Fullerton and West (2009)'s finding that peer effects are larger when defining the peers with whom an individual spends most of the time interacting.

In Columns 3 through 5 in Table 4.6, we add student, homeroom teacher, and peer characteristics one by one. Student characteristics include student gender, monthly family income, and dummies for both parents living together, father having bachelor's degrees or higher, and for mother having bachelor's degree or higher. Homeroom teacher characteristics include teacher gender and dummies for teacher experience being less than five years, teacher having graduate degree, and teacher graduating from teacher's school. Peer characteristics are leave-out means of student characteristics. The coefficients of interest are stable with the addition of those characteristics in Columns 3 through 5. When we use constant sample restricted to the observations in Column 5, we find much more stable effects (that is, minimum and maximum effects being 0.724 and 0.782.), reassuring the random assignment of peers.

We also check whether the peer effects are driven by decrease in student's height, which is implausible, as Cohen-Cole and Fletcher (2008*a*) show social network effects on height, headaches, and acnes to argue that Christakis and Fowler (2007)'s methodology is flawed. We run the same regression as Table 4.6 with different dependent variables, namely student's height and weight, separately.

Table 4.7: Are Peer Effects Driven by Height?

	OLS	2SLS			
	(1)	(2)	(3)	(4)	(5)
A. Effect on Height					
Avg. Peers' BMI	-0.569^{**}	0.176	0.141	-0.071	-0.244
	(0.271)	(0.758)	(0.907)	(0.848)	(0.941)
Own # of Siblings		-0.436^{***}	-0.219	-0.216	-0.223
		(0.159)	(0.154)	(0.154)	(0.162)
B. Effect on Weight					
Avg. Peers' BMI	-2.216^{***}	2.434^{***}	2.233^{***}	2.088^{***}	2.117^{***}
	(0.794)	(0.422)	(0.805)	(0.671)	(0.726)
Own # of Siblings		-0.833^{***}	-0.536^{**}	-0.535^{**}	-0.585^{**}
		(0.258)	(0.246)	(0.246)	(0.258)
Observations	3,909	3,909	3,630	3,630	3,464
1st Stage F-stat		7.52	7.32	9.82	8.28
School FEs	Yes	Yes	Yes	Yes	Yes
Student Ctrl			Yes	Yes	Yes
Peer Ctrl				Yes	Yes
Teacher Ctrl					Yes

Notes: Each column represents a separate regression. Dependent variables in Panels A and B are a student's height and weight, and independent variable is mean peers' BMI, which is insturmented with average peers' number of siblings in the 2SLS model. Standard errors in parentheses are clustered at school level.
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.7 clearly shows that our peer effects are driven by changes in student's weight. We also run 2SLS regression of own height on peers height using peers number of siblings as instrumental variable, but find no effects. The same regression with weight shows that a one kilogram increase in peers weight increases own weight significantly by 0.77 kilogram. Therefore, we conclude our peer effects are not driven by implausible decrease in height.

4.4.2 Effects on Obesity

Next, we examine the effect of peers' BMI on the likelihoods of student's being overweight and obese. We use the definition of adolescents being obese or overweight from 2007 WHO Reference Chart for Boys and Girls, which varies by gender and age.[37]

In Table 4.8, the dependent variable in Column 1 is dummy indicating a student's BMI being larger than or equal to that for overweight (i.e. 21.1 for boys of 13 years and 4 months, and 22.1 for girls of the same age). Column 2 is for the likelihood of obesity (i.e. BMI over 25.2 for boys and 26.6 for girls of 13 years and 4 months). Columns 1 and 2 show a one unit increase in peers' BMI increases the likelihood of student's being overweight by 10.4 percent but does not influence the likelihood of obesity.

4.4.3 Peer Effects over Time

We examine the effects of seventh grade peers' BMI over time to see if peer effects fade out as time goes by. We are able to isolate seventh grade peer effects from the later year's peer effects because students are again randomly assigned to a new classroom at the beginning of the new academic year. In Table 4.9, we regress student's weight outcomes in eighth or ninth grades on peers' BMI in seventh grade controlling for seventh grade school fixed effects.

Columns 1 and 2 show the peer effects on own BMI fade out quickly with around one third of the effects disappearing each year, and the seventh grade peer effects become

[37] Because GEPS2012 does not include student's age, we estimate the students' average age, using the fact that elementary school students in South Korea are admitted for an academic year if they turn seven during the calendar year (Korea Legislation Research Institute, 2014). For example, students born January 1st through December 31st in 1999 started their elementary school in 2006 (i.e. students born December 31st are age of seven as of December 31st in 2006, so they go to elementary school in 2006.). Assuming students were born evenly in 1999, the average age as of July 1st in 2012 will be exactly 13. Using Seoul Education Longitudinal Study of 2010 (SELS2010), which is different data set from our data and surveyed 4, 7, and 10th grade students in Seoul in 2010, we find the average age of seventh graders as of July 1st is 12.9. Because, in GEPS2012, seventh graders were surveyed in November, we assume their ages are 13 years and 4 months.

Table 4.8: Peer Effects on the Likelihood of Overweight or Obesity

	Dep. Var = Dummy for being	
	Overweight	Obese
Avg. Peers' BMI	0.108^{***}	0.032
	(0.026)	(0.022)
Own Number of Siblings	-0.034^{***}	-0.007^{*}
	(0.008)	(0.004)
Observations	3,909	3,909
R^2	0.013	0.012
1st Stage F-stat	7.54	7.54

Notes: Each column represents a separate regression, controlling for school fixed effects. Dependent variables in Columns 1 and 2 are dummies for 7th grade student's being overweight and being obese, respectively. 7th grade peers' BMI is instrumented with 7th grade peers' number of siblings. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.9: Peer Effects over Time

	BMI		Overweight		Obesity	
	8th	9th	8th	9th	8th	9th
7th Grd Avg. Peers' BMI	0.514^{*}	0.149	0.143^{***}	0.073^{*}	0.013	-0.013
	(0.273)	(0.294)	(0.044)	(0.042)	(0.025)	(0.022)
Own Number of Siblings	-0.197^{**}	-0.190^{**}	-0.020^{***}	-0.022^{***}	-0.009^{**}	-0.008^{*}
	(0.078)	(0.077)	(0.008)	(0.008)	(0.004)	(0.004)
Observations	3,556	3,467	3,556	3,467	3,556	3,467
R^2	0.021	0.043	0.001	0.034	0.018	0.036
1st Stage F-stat	7.94	7.28	7.94	7.28	7.94	7.28

Notes: Each column represents a separate regression, controlling for 7th grade school fixed effects. Dependent variables are own BMI in 8th and 9th grades, the likelihood of being overweight in 8th and 9th grades, and the likelihood of being obsese in 8th and 9th grades, respectively. Independent variable is mean peers' BMI in 7th grade, which is insturmented with average peers' number of siblings in the 2SLS model. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

insignificant in two years of the exposure to the peers. However, the effect on own BMI one year later is substantial (0.514). On top of that, we find persistent effects of seventh grade peers' BMI on the likelihood of being overweight, suggesting the importance of peers in adolescent's BMI status.

4.5 Mechanisms

GEPS2012 includes various responses on student's diet and time use which can be used to study the mechanisms behind the peer effects on BMI. In Table 4.10, we use Equation 4.2 with different dependent variables for each column. In Columns 1 through 4, we use dummy variables for eating breakfast five days or more per week, being satisfied by school lunch, eating dinner together with family five days or more per week, and not working out at all excluding P.E. classes. Dependent variables in Columns 5 through 7 are minutes spent per day on playing with friends after school, on studying, and on using a computer.

Columns 1 through 3 show that peers' BMI does not seem to influence student's daily meals. The results may be due to the responses themselves being not quite relevant; admittedly, eating dinner together with or without family could not affect student's BMI much. Nevertheless, Columns 1 and 2 indicate that peers would not affect student's number of having meals or appetite. Columns 4 through 6 show that peers' BMI does not affect the student's time use in indoor or outdoor activities. The dependent variable in Column 6 is the sum of minutes spent per day on doing homework, reading books, and self-study. Though we regress time use in those three study related activities separately, we do not find any significant effects. These results are in line with Yakusheva, Kapinos and Eisenberg (2014)'s finding that their peer effect on weight gain is not driven by peer's exercise habits or eating disorder symptoms.

Column 7 shows that a one unit increase in peers' BMI decreases student's playing time with their friends after school by 14 minutes per day. This result, combined with

Table 4.10: Mechanisms

Dep. Var. =	Dummy for				Minutes Spent per Day for		
	Breakfast 5+ days	Lunch Delicious	Dinner Family	No Exercise	Study	Computer	Play Outside
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Avg. Peers' BMI	0.052	−0.038	0.026	−0.010	10.809	−7.966	−14.325*
	(0.054)	(0.093)	(0.050)	(0.050)	(25.806)	(8.992)	(7.627)
Number of Siblings	−0.032**	−0.013	−0.016*	−0.020***	−7.739***	−1.309	3.682**
	(0.014)	(0.009)	(0.008)	(0.007)	(2.528)	(1.723)	(1.821)
Observations	3,881	3,888	3,698	3,891	3,791	3,698	3,801
R^2	0.026	0.118	0.020	0.026	0.090	0.040	0.042
1st Stage F-stat	7.52	7.54	7.31	7.50	7.47	7.33	7.53

Notes: Each column represents a separate regression. Dependent variables in Columns 1 through 4 are dummies for eating breakfast five days or more per week, being satisfied by school lunch, eating dinner together with family five days or more per week, and not working out excluding P.E. hours. Dependent variables in Columns 5 through 7 are minutes spent per day on playing with friends after school, on doing homework, reading books, and self-suty, and on using a computer. Independent variable is 7th grade peers' standardized BMI, which is instrumented with 7th grade peers' number of siblings. Standard errors in parentheses are clustered at school level.

* $p < .10$, ** $p < .05$, *** $p < .01$

the results from Columns 4 and 5, suggests that increase in peers' BMI influences own BMI through the reduced *social* outdoor activities among friends. This interpretation can explain that effect on computer use is negative, though not significant. Because students use computer frequently to play video games with their friends online,[38] it may be that the decreased social interaction among friends reduced the time to use a computer.

4.6 Conclusion

We examine the effects of peers' BMI on middle school student's various weight outcomes. Random assignment feature of our data, together with instrumental variable strategy enables us to identify large endogenous peer effects. Our result that a one unit increase in peers' BMI increases own BMI by 0.83 units implies that policy intervention to reduce overweight or obesity will have multiplier effects by influencing both targeted subjects and their peers. Our persistent effects also add the importance of peer effects.

[38]72 percent of students in our sample answered as "frequently" or "very frequently" to the question, "How often do you use computer for internet surfing or playing games?".

5. SUMMARY AND CONCLUSIONS

This dissertation finds substantial effects of students' interactions with their teacher and peers on educational outcomes. The biggest empirical challenge in these studies is nonrandom sorting of students; parents or students have some degrees of ability to choose their teachers and peers by moving house or changing address to go to a better school or by influencing class assignment. Using unique feature in secondary education in South Korea, random assignment of students into physical homeroom classroom, we show the causal links between the interactions and student's educational outcomes.

In the first paper, we show short-term effects of teacher student gender interactions on student's academic achievement. We find that switching from a male to a female teacher increases female student's test score by 10 percent of a standard deviation, relative to male student. This effect is similar to Dee (2007)'s and Carrell, Page and West (2010)'s results. Combining these similarities, the evidence on South Korea's different attitudes towards gender equality (Brandt, 2011), and the random assignment nature of our approach, we are sure that our finding reflects genuine effects that are not necessarily due to the environment being studied. Also, this effect is similar to one standard deviation increase in teacher quality in magnitude (Chetty, Friedman and Rockoff, 2014). Considering Rockoff (2004)'s finding that many observable characteristics of teacher are not correlated with teacher quality, the effects of teacher student gender interactions are important source to improve education production function.

In addition, we find the teacher-student gender matching effects persist even four years after the exposure to the teacher, which is not usual in education literature. We show our lasting effects are not driven by attrition of students, consecutive matching of female students to female teachers, or consecutive status in higher ability group. Instead, our persistent effects are due to student's choice on high school and STEM related outcomes. These findings shed light on the importance of teacher-student gender matches in policies to close the gender gap, especially in STEM fields.

In the third paper, we show the effects of middle school classroom peers' BMI on various weight outcomes. Using random assignment feature of our data and instrumental variable strategy, we identify large endogenous peer effects. Our finding that a one unit increase in peers' BMI increases own BMI by 0.83 units implies that the effects of intervention in classroom environment to reduce overweight or obesity will be multiplied by its influence on their peers as well as targeted subjects. Our somewhat persistent effects also add the importance of the peer effects in policy perspective.

In conclusion, student's interactions with teacher and peers are important policy target; the effects are sizable and do not fade out quickly. Importantly, the intervention would not incur additional costs.

REFERENCES

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2015. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics*, 33(1): 63–89.

Bagley, Sarah, Jo Salmon, and David Crawford. 2006. "Family Structure and Children's Television Viewing and Physical Activity." *Medicine and Science in Sports and Exercise*, 38(5): 910–918.

Bellizzi, Mary C., and William H. Dietz. 1999. "Workshop on Childhood Obesity: Summary of the Discussion." *The American Journal of Clinical Nutrition*, 70(1): 173–175.

Bettinger, Eric P., and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review*, 95(2): 152–157.

Brandt, Mark J. 2011. "Sexism and Gender Inequality Across 57 Societies." *Psychological Science*, 22(11): 1413–1418.

Brown, Charles, and Mary Corcoran. 1997. "Sex-Based Differences in School Content and the Male-Female Wage Gap." *Journal of Labor Economics*, 15(3): 431–465.

Burkhauser, Richard V., and John Cawley. 2008. "Beyond BMI: the Value of More Accurate Measures of Fatness and Obesity in Social Science Research." *Journal of Health Economics*, 27(2): 519–529.

Byun, Sooyong, and Kyungkeun Kim. 2010. "Educational Inequality in South Korea: The Widening Socioeconomic Gap in Student Achievement." *Research in Sociology of Education*, 17: 155–182.

Canes, Brandice J., and Harvey S. Rosen. 1995. "Following in Her Footsteps? Faculty Gender Composition and Women's Choices of College Majors." *Industrial and Labor*

Relations Review, 48(3): 486–504.

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth. 2015. "The Effect of Schooling on Cognitive Skills." *Review of Economics and Statistics*, 97(3): 533–547.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409–432.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics*, 125(3): 1101–1144.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2011. "Is Poor Fitness Contagious?: Evidence from Randomly Assigned Friends." *Journal of Public Economics*, 95(7): 657–663.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics*, 27(3): 439–464.

Chang, Virginia W., and Diane S. Lauderdale. 2005. "Income Disparities in Body Mass Index and Obesity in the United States, 1971-2002." *Archives of Internal Medicine*, 165(18): 2122–2128.

Chen, Alex Y., and José J. Escarce. 2010. "Family Structure and Childhood Obesity, Early Childhood Longitudinal Study – Kindergarten Cohort." *Preventing Chronic Disease*, 7(3): 1–8.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Cho, Insook. 2012. "The Effect of Teacher Student Gender Matching: Evidence from

OECD Countries." *Economics of Education Review*, 31(3): 54–67.

Christakis, Nicholas A., and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine*, 357(4): 370–379.

Cohen-Cole, Ethan, and Jason M. Fletcher. 2008*a*. "Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analysis." *British Medical Journal*, 337(a2533): 1–5.

Cohen-Cole, Ethan, and Jason M. Fletcher. 2008*b*. "Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic." *Journal of Health Economics*, 27(5): 1382–1387.

Corbett, Christianne, and Catherine Hill. 2015. *Solving the Equation: The Variables for Women's Success in Engineering and Computing.* DC: AAUW.

Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources*, 42(3): 528–554.

de Oliveira Meller, Fernanda, M. C. F. Assunção, A. A. Schäfer, C. L. de Mola, A. J. D. Barros, D. L. Dahly, and F. C. Barros. 2015. "The Influence of Birth Order and Number of Siblings on Adolescent Body Composition: Evidence from a Brazilian Birth Cohort Study." *British Journal of Nutrition*, 114(01): 118–125.

Epple, Dennis, and Richard Romano. 2011. "Peer Effects in Education: A Survey of the Theory and Evidence." *Handbook of Social Economics*, 1(11): 1053–1163.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review*, 104(8): 2567–2591.

Finkelstein, Eric A., Ian C. Fiebelkorn, and Guijing Wang. 2003. "National Medical Spending Attributable to Overweight and Obesity: How Much, and Who's Paying?" *Health Affairs*, 219–226.

Fowler, James H., and Nicholas A. Christakis. 2008. "Estimating Peer Effects on Health in Social Networks: A Response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, Pais." *Journal of Health Economics*, 27(5): 1400–1405.

Good, Phillip I. 2006. *Resampling Methods: A Practical Guide to Data Analysis.* 3rd ed. Boston: Birkhäuser.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics*, 1(4): 34–68.

Haugaard, Line K., Teresa A. Ajslev, Esther Zimmermann, Lars Ängquist, and Thorkild I. A. Sørensen. 2013. "Being an Only or Last-born Child Increases Later Risk of Obesity." *PLOS ONE*, 8(2): 1–9.

Hesketh, Kylie, David Crawford, Jo Salmon, Michelle Jackson, and Karen Campbell. 2007. "Associations between Family Circumstance and Weight Status of Australian Children." *International Journal of Pediatric Obesity*, 2(2): 86–96.

Hoffmann, Florian, and Philip Oreopoulos. 2009. "A Professor like Me: the Influence of Instructor Gender on College Achievement." *Journal of Human Resources*, 44(2): 479–494.

Holmlund, Helena, and Krister Sund. 2008. "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics*, 15(1): 37–53.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." *NBER Working Paper No. 7867*.

Jackson, C. Kirabo. 2012. "Single-sex Schools, Student Achievement, and Course Selection: Evidence from Rule-based Student Assignments in Trinidad and Tobago." *Journal of Public Economics*, 96(1): 173–187.

Jacob, Brian A., Lars Lefgren, and David P. Sims. 2010. "The Persistence of Teacher-induced Learning." *Journal of Human resources*, 45(4): 915–943.

Jacoby, Ann, D. G. Altman, Judith Cook, W. W. Holland, and Elliottm A. 1975. "Influence of Some Social and Environmental Factors on the Nutrient Intake and Nutritional Status of Schoolchildren." *British Journal of Preventive and Social Medicine*, 29(2): 116–120.

Jones, Susanne M., and Kathryn Dindia. 2004. "A Meta-analytic Perspective on Sex Equity in the Classroom." *Review of Educational Research*, 74(4): 443–471.

Kang, Changhui. 2007. "Classroom Peer Effects and Academic Achievement: Quasi-randomization Evidence from South Korea." *Journal of Urban Economics*, 61(3): 458–495.

Korea Legislation Research Institute. 2011. "Enforcement Decree of the Elementary and Secondary Eduation Act, Article 68 (Methods for Entering Middle Schools)." *Presidential Decree No.22712*.

Korea Legislation Research Institute. 2014. "Elementary and Secondary Education Act, Article 13 (Obligation to Send Children to School)." *Act No.12338*.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, 3(2): 1–33.

Lavy, Victor, and Edith Sand. 2015. "On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers Stereotypical Biases." *NBER Working Paper No. 20909*.

Lee, Soohyung, Lesley J. Turner, Seokjin Woo, and Kyunghee Kim. 2014. "All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement." *NBER Working Paper No. 20722*.

Lehmann, Erich L., and Joseph P. Romano. 2005. *Testing Statistical Hypotheses (Springer Texts in Statistics).* . 3rd ed., Springer.

Lim, Jaegeum, and Jonathan Meer. forthcoming. "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea." *Journal of Human Resources*.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

Mora, Toni, and Joan Gil. 2013. "Peer Effects in Adolescent BMI: Evidence from Spain." *Health Economics*, 22(5): 501–516.

Muralidharan, Karthik, and Ketki Sheth. 2016. "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers." *Journal of Human Resources*, 51(2): 269–297.

OECD. 2010. *Obesity and the Economics of Prevention.* OECD Publishing.

OECD. 2015*a*. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence.* OECD Publishing.

OECD. 2015*b*. *Health at a Glance 2015.* OECD Publishing.

Ogden, Cynthia L., Margaret D. Carroll, Brian K. Kit, and Katherine M. Flegal. 2014. "Prevalence of Childhood and Adult Obesity in the United States, 2011-2012." *Journal of the American Medical Association*, 311(8): 806–814.

Paredes, Valentina. 2014. "A Teacher like Me or a Student like Me? Role Model versus Teacher Bias Effect." *Economics of Education Review*, 39(2): 38–49.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi. 2013. "Causal Effects of Single-sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools." *Demography*, 50(2): 447–469.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2): 247–252.

Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial & Labor Relations Review*, 48(3): 515–530.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, 125(1): 175–214.

Seoul Metropolitan Office of Education. 2011. "Guidelines on Seoul High School Curriculum." *SMOE Guidelines No.2012-14*.

Seoul Metropolitan Office of Education. 2012. "Announcement on High School Admissions Plan for 2013 School Year." *SMOE Announcement No.2012-51*.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology*, 35(1): 4–28.

Trogdon, Justin G., James Nonnemaker, and Joanne Pais. 2008. "Peer Effects in Adolescent Overweight." *Journal of Health Economics*, 27(5): 1388–1399.

Whitmore, Diane. 2005. "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *American Economic Review*, 95(2): 199–203.

Yakusheva, Olga, Kandice A. Kapinos, and Daniel Eisenberg. 2014. "Estimating Heterogeneous and Hierarchical Peer Effects on Body Weight Using Roommate Assignments as a Natural Experiment." *Journal of Human Resources*, 49(1): 234–261.

Yakusheva, Olga, Kandice A. Kapinos, and Marianne Weiss. 2011. "Peer Effects and the Freshman 15: Evidence from a Natural Experiment." *Economics & Human Biology*, 9(2): 119–132.